

République Algérienne Démocratique et Populaire
Université Abderrahmane MIRA de Béjaïa
Faculté des Sciences Exactes

Département de Recherche Opérationnelle



جامعة بجاية
Tasdawit n Bgayet
Université de Béjaïa



Mémoire présenté pour l'obtention du diplôme de Master
en Mathématiques Appliquées

Spécialité : Sciences de Données et Aide à la Décision

**Approche Data-Driven pour l'optimisation des
processus métiers chez Ovitale**

Présenté par :
MADANI Katia
HAMIDOUCHE Thinhinane

Sous la direction de : Mr. LAOUAR Abdelhek

Défendu le 29/06/2025, devant le jury composé de :

M ^r BRAHMI Belkacem	Président	MCA	Université de Béjaïa
M ^r ASLI Larbi	Examineur	MCA	Université de Béjaïa
M ^{elle} TETAH Ikram	Examinatrice	Doctorante	Université de Béjaïa

Année Universitaire 2024-2025

Remerciements

Avant tout, nous exprimons notre profonde gratitude à ALLAH, le Tout-Puissant, de nous avoir accordé la force, la patience et le courage nécessaires pour mener à bien ce travail.

Nous tenons à adresser nos sincères remerciements à Mr.LAOUAR, notre encadrant, pour sa disponibilité et son accompagnement rigoureux tout au long de ce travail, ainsi qu'à Mr.ASLI, chef de notre département, pour son soutien constant et ses efforts continus en faveur de la réussite des étudiants.

Nous adressons aussi nos plus vifs remerciements à Mr.KHALDI Yacine, dont l'engagement et le soutien ont été déterminants dans l'aboutissement de ce travail. Sa disponibilité, ses conseils avisés et ses encouragements constants nous ont profondément marqués. À vous dire vrai, les mots nous manquent pour exprimer toute notre reconnaissance.

Nous exprimons également notre profonde reconnaissance à l'ensemble de nos enseignants de l'Université de Béjaïa, ainsi qu'à tous nos professeurs depuis l'école primaire, pour la richesse des savoirs transmis et les valeurs qu'ils nous ont inculquées tout au long de notre parcours académique.

Nos remerciements les plus respectueux s'adressent également à l'entreprise OVITALE, qui nous a offert l'opportunité d'effectuer notre stage au sein de ses équipes. Nous remercions tout particulièrement Mr.HAMIDOUCHE, Mr.AIT OUALI, Mr.BOUSKINE et Mr.HEMMAM pour leur accueil chaleureux, leur disponibilité et leur encadrement bienveillant durant cette expérience professionnelle enrichissante.

Enfin, nous tenons à exprimer notre gratitude la plus profonde à nos chères familles, pour leur soutien inconditionnel, leur amour, leurs sacrifices et leurs encouragements tout au long de notre parcours. Nous remercions également toutes les personnes qui ont contribué, de près ou de loin, à la réalisation de ce projet.

Dédicace

Je dédie ce travail :

À mes parents

Cette réussite, je la dois à vos sacrifices, à votre soutien indéfectible et à vos prières. Merci pour votre amour et votre foi en moi. Les mots ne suffisent pas pour vous remercier. J'espère de tout cœur que vous êtes fiers, car ce succès est avant tout le vôtre.

À mes grands-parents

À mon cher grand-père, ta présence m'apporte tant de bonheur et de réconfort. Fière d'être ta petite-fille, je prie qu'Allah t'accorde une longue vie en santé.

À ma chère grand-mère, ton absence pèse lourdement en ce moment si particulier de ma vie. J'aurais tant aimé pouvoir partager cette réussite avec toi. Je te dédie ce travail avec tout mon amour, *Allah yarhemak*.

À mon oncle Yacine

Tu occupes une place toute particulière dans mon cœur. Ta présence et ta vision des choses ont guidé mes pas et ont profondément influencé mon parcours. Si j'ai atteint cette réussite aujourd'hui, c'est en grande partie grâce à toi. Merci du fond du cœur.

À Thinhinane, ma précieuse binôme

Merci pour ton engagement, ton esprit d'équipe et ton soutien tout au long de ce projet. Cette réussite est le fruit de notre belle collaboration.

À tous ceux qui m'ont toujours encouragé et étaient à mes côtés.
Je vous dis merci.

Katia

Dédicace

Je dédie ce travail :

À mes parents

Mes piliers, mes forces, ceux qui ont toujours porté mes rêves avant moi. Leur courage, leurs sacrifices et leur patience sont la vraie raison pour laquelle je suis là aujourd'hui. Aucun mot ne pourra jamais suffire pour leur dire merci.

À mes frères et sœurs, et à toute ma famille

À mes deux grands-mères, Yemma Wawa et Yemma A Zizou, à tous mes proches, et à toutes ces personnes vraies qui m'aiment sincèrement, sans masque ni intérêt, avec un cœur pur. Merci d'avoir toujours été là.

À Kamilia, ma seule amie, ma sœur de cœur

Toujours présente, dans le silence comme dans les épreuves. Merci d'exister.

À mes enseignants

Du primaire jusqu'à l'université, vous avez semé en moi une part de ce que je suis aujourd'hui. Merci pour votre savoir, votre patience et votre dévouement.

À Katia, mon binôme

Avec qui j'ai partagé cette belle aventure. Merci pour ton sérieux, ton engagement et ta sincérité.

À la petite Thinhinane

Celle qui, à 5 ans, faisait ses premiers pas à l'école, curieuse, pleine de rêves. Aujourd'hui, je lui dis doucement : « Tu peux être fière, petite fille. Le chemin a été long, mais tu l'as fait. Et ce n'est que le début. »

À tous ceux qui m'ont toujours encouragée et soutenue avec bienveillance.
Je vous dis simplement et sincèrement : merci.

Thinhinane

Table des matières

Table des matières	I
Table des figures	IV
Liste des tableaux	VI
Liste des abréviations et notations	VII
Introduction générale	1
1 l'industrie à l'ère du Big Data	3
Introduction	3
1.1 L'évolution industrielle	3
1.2 Une industrie 4.0 fondée sur l'économie de la connaissance	5
1.3 L'industrie 4.0 - Révolution numérique de la production	6
1.3.1 Caractéristiques fondamentales	7
1.3.2 Technologies clés	8
1.3.3 Différences avec les modèles traditionnels	8
1.3.4 Défis et perspectives	9
1.4 Big Data et Cloud Computing au cœur de l'industrie 4.0	9
1.4.1 Cloud Computing	10
1.4.2 Big Data	10
Conclusion	11
2 Data Science for Business	12
Introduction	12
2.1 Data Science	12
2.1.1 Mathématiques pour la Data Science	13
2.1.2 Data Visualisation	14
2.2 Machine learning	15
2.2.1 Démarche méthodologique d'un projet de Machine Learning (ML)	15
2.2.2 Métriques d'évaluation	18
2.2.3 Défis majeurs du ML	19
2.3 Business Intelligence	20
2.3.1 Méthodologie de la Business Intelligence (BI)	20
2.3.2 Architecture de stockage	21
2.3.3 Modélisation des données en BI	22

2.4	Synergie entre BI et Data Science	23
2.5	Optimisation de la chaîne d’approvisionnement	23
2.5.1	Processus d’optimisation de la chaîne d’approvisionnement	24
	Conclusion	25
3	Méthodologie Data-Driven	27
	Introduction	27
3.1	Le concept de Data-Driven	27
3.1.1	Dans la littérature professionnelle	28
3.1.2	Dans la littérature organisationnelle	28
3.2	Définition d’une organisation data-driven	28
3.3	Les fondations d’une organisation data-driven	30
3.4	Processus de transformation vers une organisation data-driven	30
3.4.1	Analyse des besoins	30
3.4.2	Collecte des données	31
3.4.3	Qualité des données	32
3.4.4	Accès aux données	32
3.4.5	Organisation des équipes analytiques	33
3.4.6	Du reporting à l’analyse	34
3.4.7	Analyse de données	34
3.4.8	Les indicateurs clés de performance (KPIs)	35
3.4.9	Data culture	36
3.4.10	Les outils clés pour devenir une entreprise data-driven	38
3.4.10.1	Plateformes de BI	38
3.4.10.2	Outils d’analyse prédictive	39
3.4.10.3	CRM intégrés	40
3.4.10.4	Outils de collecte de données	40
3.4.10.5	Outils de gestion des données	41
3.4.10.6	Plateformes de Marketing Automation	41
3.5	Success Stories : L’impact du data-driven sur les entreprises populaires	42
3.5.1	Amazon	42
3.5.2	Starbucks	43
3.5.3	IBM	44
3.5.4	Netflix	45
	Conclusion	46
4	Étude de cas : Optimisation des processus métiers chez Ovitale	47
	Introduction	47
4.1	Présentation de l’entreprise Ovitale	47
4.2	Analyse de besoins	49
4.2.1	Diagnostic initial	49
4.2.2	Définition des objectifs stratégiques	49
4.3	Collecte des données	50
4.3.1	Données internes	50
4.3.2	Données externes (contextuelles)	50
4.4	Traitement des données pour l’analyse	51

4.5	Analyse descriptive	52
4.5.1	Analyse descriptive des ventes	53
4.5.2	Analyse descriptive de la production	57
4.5.3	Suivi de stock	61
4.5.4	Analyse descriptive du l'approvisionnement	63
4.6	Analyse diagnostique	67
4.7	Analyse prédictive	70
4.7.1	Préparation des données	70
4.7.2	Séparation des données	71
4.7.3	Choix et évaluation des modèles de prévision	71
4.7.4	Résultats obtenus	72
4.7.5	Visualisation interactive des prévisions	76
4.8	Analyse prescriptive	76
4.8.1	Planification de la production	77
4.8.2	Visualisation interactive de la planification de la production	79
4.8.3	Planification de l'approvisionnement et gestion des stocks	79
4.8.4	Visualisation interactive de la planification d'approvisionnement et de la gestion des stocks	83
4.8.5	Recommandation des fournisseurs	83
4.8.6	Visualisation interactive des recommandations de fournisseurs	84
4.9	Pipeline Analytique : De Python à Power BI	85
	Conclusion	86
	Conclusion générale	87
	Bibliographie	89
	Webographie	91

Table des figures

1.1	Évolution de l'industrie.	4
1.2	Structure de la production technologique dans l'économie de la connaissance et l'industrie 4.0.	5
1.3	L'évolution de la société.	6
1.4	L'écosystème de l'industrie 4.0	7
1.5	Rôle du Big Data et du Cloud dans l'Industrie 4.0.	9
2.1	Compétences et domaines clés en Data Science.	13
2.2	Cycle de vie des données.	21
3.1	Organisation data-driven.	29
3.2	Les fondations d'une organisation data-driven.	30
3.3	Impact du leadership sur l'adoption d'une culture orientée données.	37
4.1	Logo d'Ovitale.	48
4.2	Les trois formats de bouteilles produits par Ovitale.	49
4.3	Schéma en constellation 1.	52
4.4	Suivi des ventes.	53
4.5	Évolution mensuelle et saisonnière des ventes (2020–2024).	54
4.6	Répartition des ventes par format de Produit fini (PF) (2020–2024).	55
4.7	Répartition des quantités vendues par client (Top 5) (2020–2024).	56
4.8	Répartition géographique des ventes par wilaya (Top 5) (2020-2024).	57
4.9	Suivi de la production.	58
4.10	Évolution mensuelle de la production (2020–2024).	59
4.11	Évolution mensuelle de la consommation et des pertes de matières premières (MP) (2020–2024).	60
4.12	Consommation et pertes cumulées des MP (2020–2024).	61
4.13	Suivi de stock – Quantité.	62
4.14	Suivi de stock – Montant	62
4.15	Suivi de l'approvisionnement.	63
4.16	Évolution mensuelle de la quantité et du montant des MP achetées (2020–2024).	64
4.17	Quantité achetée et montant cumulé des MP (2020–2024).	65
4.18	Segmentation des fournisseurs par K-means.	66
4.19	Segmentation des fournisseurs et quantités achetées par classe.	66
4.20	Suivi des ventes selon les périodes spéciales et conditions climatiques.	67
4.21	Évolution mensuelle des ventes et de la production (2020–2024).	68

4.22	Évolution des ventes en fonction de la température et de l'humidité moyennes (2020–2024).	68
4.23	Évolution des ventes moyennes selon les périodes spéciales (2020–2024). . . .	69
4.24	Jeu de données de ventes.	71
4.25	Exemple d'un jeu de données d'entraînement de l'article 1002.	71
4.26	Prévisions de la demande pour l'article 1002	73
4.27	Prévisions de la demande pour l'article 1003	74
4.28	Prévisions de la demande pour l'article 1003	75
4.29	Suivi de la demande future.	76
4.30	Aperçu hebdomadaire du plan de production optimal pour l'article 1002.	78
4.31	Planification de la production.	79
4.32	Nomenclature des PF.	80
4.33	Aperçu hebdomadaire de la planification Planification des besoins en matières (MRP) des MP.	82
4.34	Aperçu des coûts d'approvisionnement des MP.	82
4.35	Planification de l'Approvisionnement.	83
4.36	Recommandation des fournisseurs et répartition des quantités par MP.	84
4.37	Recommandation des fournisseurs.	84
4.38	Schéma en constellation 2.	85

Liste des tableaux

1.1	Technologies et concepts clés de l'industrie 4.0	8
2.1	Métriques d'évaluation pour les modèles de classification.	18
2.2	Métriques d'évaluation pour les modèles de régression.	19
3.1	Les quatre principaux types d'analyse de données	35
3.2	Exemples d'objectifs smart en industrie	36
3.3	Top 3 des logiciels de BI en 2025.	39
3.4	Les 3 meilleurs logiciels d'analyse prédictive.	39
3.5	Exemples d'outils systèmes de gestion de la relation client (CRM) intégrés. . .	40
3.6	Exemples de services de collecte de données en 2025.	41
3.7	Exemples de plateformes de Marketing Automation.	42
3.8	Stratégies Data-Driven d'Amazon.	43
3.9	Stratégies Data-Driven de Starbucks.	44
3.10	Stratégies Data-Driven d'IBM.	45
3.11	Stratégies Data-Driven de Netflix.	46
4.1	Historique d'Ovitale	48
4.2	Sources internes de données utilisées dans le projet.	50
4.3	Sources et outils de collecte des données externes.	51
4.4	Évolution des indicateurs clés de performance (KPIs) des ventes (2020–2024). .	53
4.5	Évolution des KPIs de production (2020–2024).	58
4.6	Évolution des KPIs d'approvisionnement (2020–2024).	63
4.7	Performances des modèles de ML pour l'article 1002.	73
4.8	Performances des modèles de ML pour l'article 1003.	74
4.9	Performances des modèles pour l'article 1006	75
4.10	Synthèse des résultats d'optimisation pour les différents formats de production .	78
4.11	Synthèse de la planification de l'approvisionnement et de la gestion des stocks en MP.	82

Liste des abréviations et notations

IA	Intelligence Artificielle
IoT	Internet des Objets
BI	Business Intelligence
ML	Machine Learning
CPS	systèmes cyber-physiques
IoS	internet des services
PME	petites et moyennes entreprises
ES	Systèmes embarqués
RFID	Identification par radiofréquence
AM	Fabrication additive
PLM	Gestion du cycle de vie des produits
HCI	Interaction Homme-Machine
AR	Réalité augmentée
VR	Réalité virtuelle
MR	Réalité mixte
BCG	Boston Consulting Group
API	Interface de Programmation Applicative
KPIs	indicateurs clés de performance
HiPPOs	Highest Paid Person's Opinion
FTC	Federal Trade Commission
CRM	systèmes de gestion de la relation client
CA HT	chiffre d'affaires hors taxe
CA TTC	Chiffre d'affaires toutes taxes comprises
CA	chiffre d'affaires
MP	matières premières
PF	Produit fini
ARIMA	Autoregressive Integrated Moving Average
SARIMA	Seasonal Autoregressive Integrated Moving Average
RMSE	Root Mean Squared Error

MAE	Mean Absolute Error
MSE	Mean Squared Error
SVR	Support Vector Regression
KNN	K-Nearest Neighbors
XGBoost	Extreme Gradient Boosting
LightGBM	Light Gradient Boosting Machine
AdaBoost	Adaptive Boosting
Bagging	Bootstrap Aggregating
MLPRegressor	Multi-Layer Perceptron Regressor
DA	Dinar Algérien
EOQ	Quantité Économique à Commander
EPQ	Quantité Économique à Produire
MRP	Planification des besoins en matières
BB	Besoins Bruts
BN	Besoins Nets
PDP	Plan Directeur de Production
SP	Stock Prévisionnel
SI	Stock Initial
SS	Stock de Sécurité
ETL	Extract Transform Load
TP	vrais positifs
FP	faux positifs
TN	vrais négatifs
FN	faux négatifs
OLAP	Online Analytical Processing
DWH	Datawarehouse
PIC	Plan Industriel et Commercial
TOC	Théorie des Contraintes
VMI	Gestion des stocks par le fournisseur

Introduction générale

La transformation numérique bouleverse profondément le paysage industriel actuel. L'émergence de technologies telles que l'Internet des Objets (IoT), l'Intelligence Artificielle (IA) et les systèmes de cloud computing a engendré une explosion du big data, offrant un potentiel inédit pour optimiser les processus et innover dans divers secteurs. Dans ce contexte, les données ne sont plus de simples éléments informatifs, mais constituent un véritable actif stratégique, moteur de la compétitivité et de la performance.

Le secteur industriel connaît aujourd'hui une mutation accélérée, portée par des technologies émergentes qui redéfinissent les méthodes de production et de gestion. La collecte massive de données, rendue possible par l'IoT et la digitalisation des processus, permet aux entreprises de prendre des décisions éclairées et de mieux anticiper les évolutions du marché. Cette révolution numérique impose aux industries de repenser leurs modèles opérationnels pour rester compétitives et répondre aux exigences d'un environnement en perpétuelle évolution.

La transition d'une économie traditionnelle, fondée sur les ressources matérielles, vers une économie de la connaissance met en avant la valorisation de l'information. Dans ce nouveau paradigme, l'analyse et l'exploitation des données deviennent des leviers essentiels pour l'innovation et la prise de décision stratégique. Les entreprises qui réussissent à transformer ces données en connaissances opérationnelles se positionnent comme des leaders sur leurs marchés respectifs.

À l'ère de la transformation numérique, les entreprises industrielles adoptent progressivement de nouvelles approches fondées sur l'exploitation intelligente des données. Cette dynamique, portée par l'essor des technologies connectées et des systèmes automatisés, s'inscrit dans le contexte plus large de l'industrie 4.0, qui ouvre la voie à une production plus agile, plus efficiente et résolument axée sur les données.

Face à cette réalité, l'approche data-driven se présente comme une méthode incontournable pour exploiter pleinement le potentiel des données. En adoptant une démarche basée sur l'analyse des données, les entreprises peuvent optimiser leurs processus métiers, améliorer leur réactivité et anticiper les besoins du marché. Cette approche, en synergie avec les outils de la data science, BI et les modèles de ML, constitue un levier essentiel pour réussir la transformation digitale, notamment dans le cadre de l'industrie 4.0.

Dans ce contexte de transformation industrielle, Ovitale s'inscrit déjà dans la dynamique de l'industrie 4.0, en intégrant certaines technologies avancées dans ses processus. Toutefois, des défis persistent en matière d'optimisation, notamment au niveau de la gestion commerciale, de la production, des approvisionnements et des stocks.

Conscients de l'importance croissante des données dans le processus de décision, nous proposons d'adopter une approche data-driven pour accompagner Ovitale dans l'amélioration de ses processus métiers. Cette démarche vise à structurer et exploiter efficacement les données existantes afin de renforcer la performance opérationnelle de l'entreprise.

Nous considérons qu'une organisation véritablement pilotée par les données constitue une étape essentielle pour tirer pleinement parti des opportunités offertes par cette transition numérique. En soutenant cette évolution, nous cherchons à démontrer comment une telle stratégie peut être mise en œuvre concrètement, et quels bénéfices elle peut générer pour une entreprise algérienne comme Ovitale.

Afin de répondre à cette problématique, ce mémoire est structuré en quatre chapitres complémentaires :

- Le premier chapitre retrace l'évolution industrielle, souligne le rôle de l'économie de la connaissance dans l'industrie 4.0, et explore les technologies et implications du Big Data dans cette révolution numérique.
- Le deuxième chapitre est consacré aux fondements théoriques de la data science et de la BI, ainsi qu'aux algorithmes contribuant à l'optimisation de la chaîne logistique. Il explore les concepts clés de ces disciplines et met en lumière la synergie entre la data science et la BI.
- Le troisième chapitre développe la méthodologie de l'approche data-driven, en décrivant les étapes de mise en œuvre, les outils mobilisés et les modèles d'analyse adoptés pour accompagner la transition vers une entreprise pilotée par les données.
- Enfin, le quatrième chapitre est dédié à l'étude de cas de l'entreprise Ovitale. Il présente la démarche suivie, depuis la compréhension des besoins jusqu'à l'implémentation de tableaux de bord intelligents pour l'optimisation des processus commerciaux, de production, d'approvisionnement et de gestion des stocks.

Nous terminons par une conclusion générale qui récapitule les résultats obtenus.

1

l'industrie à l'ère du Big Data

Introduction

L'industrie contemporaine est en constante mutation, portée par les avancées technologiques et la transformation numérique. Dans ce contexte, le big data apparaît comme un catalyseur majeur de changement, bouleversant les modèles traditionnels de production et de gestion. L'émergence de l'industrie 4.0, considérée comme la quatrième révolution industrielle, repose sur une exploitation stratégique des données massives et une automatisation intelligente des processus.

Ce chapitre se propose d'établir un cadre conceptuel à cette évolution industrielle. Il débute par une rétrospective des différentes phases de l'industrialisation, de la mécanisation à l'intelligence artificielle, avant de s'intéresser à l'économie de la connaissance, vecteur essentiel du développement de l'industrie 4.0. Enfin, il met en lumière les caractéristiques clés, les technologies fondamentales et les implications de cette révolution numérique.

Ces éléments sont indispensables pour comprendre les enjeux du big data dans les environnements industriels modernes, et préparent le lecteur à l'analyse plus technique qui sera développée dans les chapitres suivants, notamment dans notre cas d'étude.

1.1 L'évolution industrielle

L'évolution industrielle a débuté dès la première révolution, à la fin du 18e siècle en Angleterre, transformant en profondeur les secteurs de la fabrication, de la technologie et de la société. Depuis lors, elle n'a cessé d'évoluer, chaque phase étant marquée par des avancées technologiques et sociales significatives (Fig4.1). Cette dynamique progressive a conduit à l'émergence de l'économie numérique moderne, avec l'hypothèse d'une cinquième révolution industrielle en perspective :

- **Industrie 1.0** : cette révolution a marqué le passage du travail manuel à la fabrication mécanisée grâce à des innovations comme la machine à vapeur, révolutionnant les transports, la production et les secteurs de l'agriculture, des textiles et de la métallurgie.
- **Industrie 2.0** : à la fin du 19e et au début du 20e siècle, cette phase a vu l'introduction de l'électricité, des chaînes de montage et l'essor des industries chimiques et pétrolières, favorisant une production de masse et une accélération de la croissance économique.
- **Industrie 3.0** : à la fin du 20e siècle, l'électronique, les ordinateurs et l'internet ont transformé l'industrie, introduisant l'automatisation et la numérisation, et donnant naissance à l'économie numérique.
- **Industrie 4.0** : actuellement en cours, cette phase se caractérise par l'automatisation avancée, l'IA et l'IoT, permettant une intégration numérique accrue dans la fabrication et la gestion, et une prise de décision plus efficace basée sur les données.
- **Industrie 5.0** : cette phase potentielle met l'accent sur la collaboration entre humains et robots, et la personnalisation dans la fabrication, soutenue par des avancées en IA, 5G et mondialisation numérique, promettant des améliorations en connectivité et en transfert de données.

L'évolution industrielle a progressivement intégré des technologies avancées, faisant des données un atout stratégique essentiel à chaque étape. L'industrie 4.0, axée sur l'automatisation, l'IA et l'IoT, souligne l'importance croissante de la collecte et de l'analyse des données. Aujourd'hui, les entreprises doivent exploiter ces données pour optimiser leurs processus et améliorer la prise de décision, marquant ainsi un tournant vers une gestion pilotée par les données pour rester compétitives [Vor+23, Özü+18].

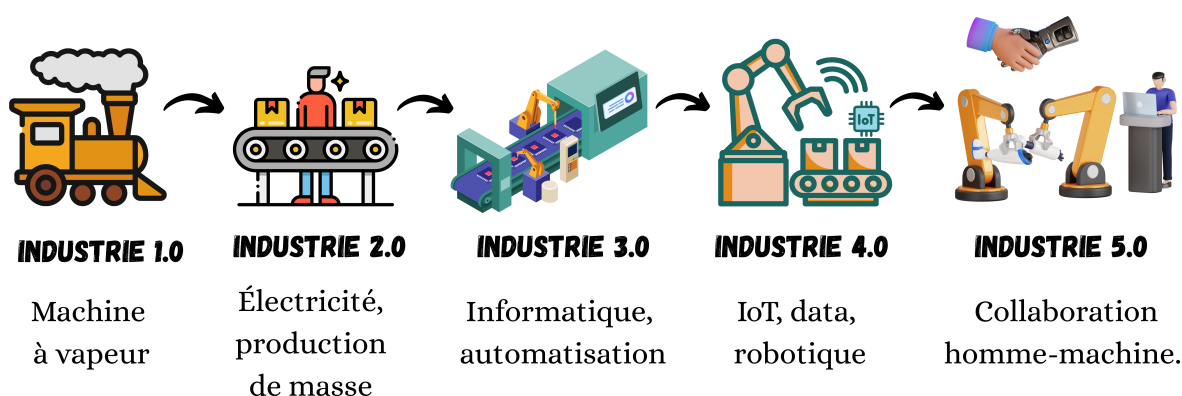


FIGURE 1.1 – Évolution de l'industrie.

1.2 Une industrie 4.0 fondée sur l'économie de la connaissance

L'économie de la connaissance (en anglais knowledge economy) désigne un modèle économique dans lequel la création, la diffusion et l'exploitation des connaissances sont les principaux moteurs de croissance et de compétitivité. Contrairement aux économies traditionnelles, basées sur les ressources naturelles ou la force de travail physique, elle repose sur l'innovation, l'éducation et les technologies de l'information.

L'économie de la connaissance crée un environnement favorable à l'émergence de nouvelles technologies, d'innovations et de compétences humaines essentielles au développement de l'industrie 4.0. Cette dernière devient un pilier central de l'économie de la connaissance, en intégrant ces technologies et innovations dans les processus industriels. Ainsi, un cercle vertueux se forme, où chaque élément nourrit et soutient l'autre, consolidant leur développement mutuel.

La structure du mode de production technologique pour l'économie de la connaissance et l'industrie 4.0 peut être représentée de manière schématique comme suit (Fig1.2) :

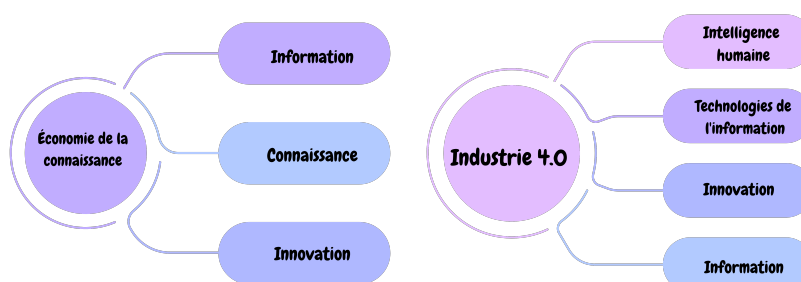


FIGURE 1.2 – Structure de la production technologique dans l'économie de la connaissance et l'industrie 4.0.

Dans ce contexte, plusieurs facteurs clés illustrent la manière dont l'économie de la connaissance soutient concrètement la transition vers l'industrie 4.0 et en maximise le potentiel (Fig1.3) :

- **Développement et transition vers l'industrie 4.0** : il est essentiel pour le développement de l'industrie 4.0, en soutenant l'intégration des technologies avancées et des processus innovants. Cette évolution nécessite une économie solide, favorisant la créativité, l'apprentissage continu et le développement du capital humain, facilitée par la gestion des connaissances et les progrès technologiques.
- **Soutien à l'innovation et performances** : il stimule la croissance des industries technologiques et l'entrepreneuriat innovant, tout en optimisant la production industrielle et l'efficacité organisationnelle, ce qui est essentiel pour la transition vers l'industrie 4.0.

- **Capital humain et innovation** : il l'importance du capital humain, de l'innovation et de l'échange d'informations pour la compétitivité des entreprises industrielles.
- **Gestion stratégique des connaissances** : des systèmes efficaces de gestion des connaissances sont essentiels pour maintenir un avantage concurrentiel, soutenus par une direction et une culture organisationnelle favorables à l'innovation.
- **Exploitation des connaissances internes et externes** : Les entreprises doivent tirer parti des connaissances internes et collaborer avec des experts externes pour maximiser le potentiel de l'industrie 4.0 [Bog18].

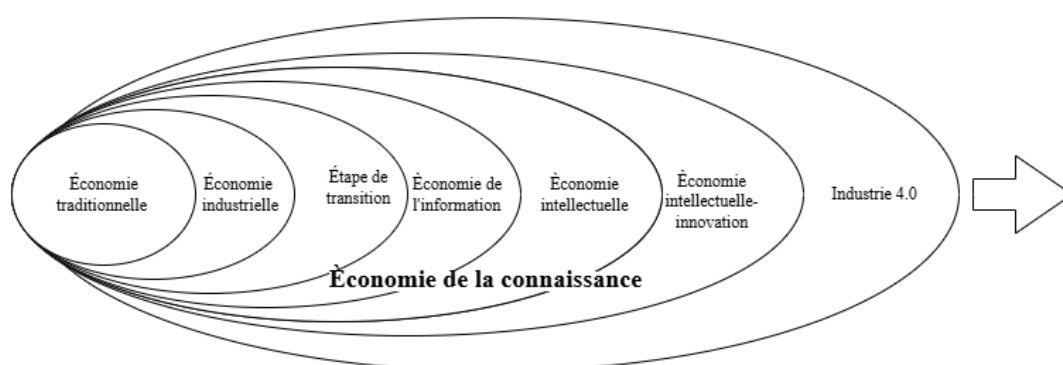


FIGURE 1.3 – L'évolution de la société.

1.3 L'industrie 4.0 - Révolution numérique de la production

Le terme industrie 4.0, apparu en 2011 à la foire de Hanovre, désigne la quatrième révolution industrielle fondée sur l'intégration du numérique dans la production. Initiée en Allemagne pour renforcer sa compétitivité face à des puissances comme la Chine, cette démarche a été rapidement intégrée à la stratégie nationale d'innovation, la Stratégie High-Tech 2020.

L'industrie 4.0 désigne la transformation numérique des chaînes de valeur grâce à des technologies comme les systèmes cyber-physiques (CPS), l'IoT, la gestion des données massives (big data) et l'IA. Elle repose sur une collaboration en temps réel entre machines, données et humains, formant un écosystème intelligent, où les équipements communiquent de manière autonome pour optimiser la production (Fig 1.4) [BN22].

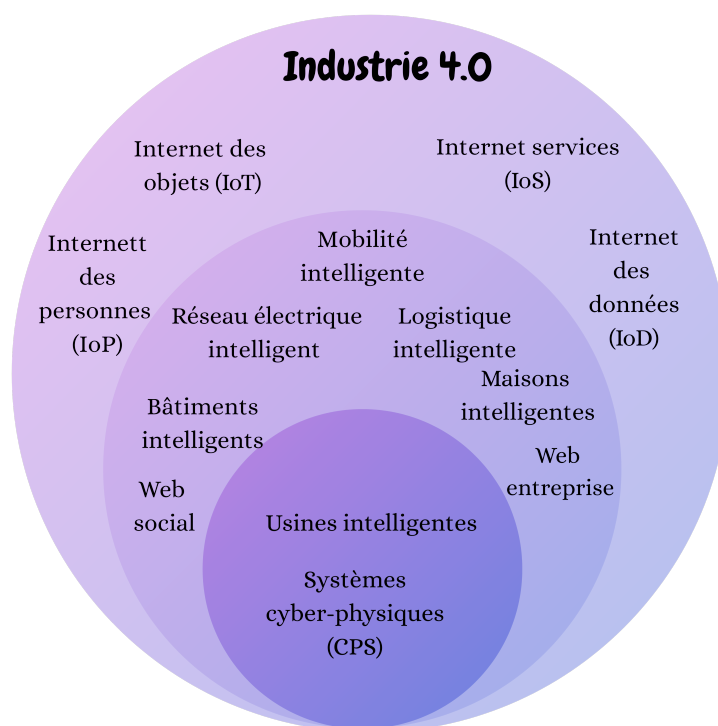


FIGURE 1.4 – L'écosystème de l'industrie 4.0

1.3.1 Caractéristiques fondamentales

Malgré la diversité des définitions, les experts s'accordent sur plusieurs caractéristiques clés de l'industrie 4.0 :

- **Interopérabilité** : capacité des machines, systèmes et personnes à communiquer et coopérer grâce aux réseaux numériques comme l'IoT et l'internet des services (IoS).
- **Virtualisation** : création de jumeaux numériques (répliques virtuelles) des installations physiques, permettant de simuler et de surveiller les opérations en temps réel.
- **Décentralisation** : autonomie des systèmes à prendre des décisions localement, sans supervision centrale, augmentant ainsi la flexibilité et la capacité à résoudre des problèmes complexes.
- **Capacité en temps réel** : aptitude des dispositifs à collecter et analyser instantanément les données pour réagir rapidement aux changements ou anomalies dans le processus de production.
- **Orientation service** : production centrée sur le client, grâce à l'analyse de ses préférences via les réseaux intelligents et les plateformes numériques.
- **Modularité** : flexibilité des systèmes industriels à être reconfigurés rapidement selon les besoins, facilitant l'adaptation aux fluctuations du marché [BN22].

1.3.2 Technologies clés

L'environnement de l'industrie 4.0 repose sur un ensemble de technologies interconnectées, formant un écosystème intelligent dont les composants clés sont illustrés dans le tableau ci-dessous (1.1) :

Groupe	Technologies et concepts
Usines intelligentes	Systèmes Cyber-Physiques (CPS) Systèmes embarqués (ES) Identification par radiofréquence (RFID) Internet des Objets (IoT) Internet des Services (IoS) Automatisation Modularisation Fabrication additive (AM) Gestion du cycle de vie des produits (PLM) Robotique Interaction Homme-Machine (HCI) Outils et modèles de simulation
Simulation et modélisation	Réalité augmentée (AR) Réalité virtuelle (VR) Réalité mixte (MR)
Numérisation et virtualisation	Cloud Computing Big Data Informatique mobile Réseaux sociaux Numérisation

TABLE 1.1 – Technologies et concepts clés de l'industrie 4.0

Ces technologies permettent de créer des usines intelligentes, capables de s'auto-organiser, de prédire les défaillances, de lancer automatiquement les processus de maintenance et de coordonner la logistique de façon autonome [BN22].

1.3.3 Différences avec les modèles traditionnels

Le concept d'industrie 4.0 se distingue fondamentalement des modèles industriels traditionnels, avec des différences marquées dans plusieurs domaines clés :

- **Produits** : personnalisation accrue avec une production optimisée et à faible coût.
- **Processus** : solutions en réseau réduisant les barrières technologiques et favorisant les petites et moyennes entreprises (PME).

- **Modèles économiques** : nouvelles opportunités pour les petites entreprises sur les marchés globaux.
- **Concurrence** : accès élargi aux marchés mondiaux.
- **Compétences** : évolution vers des méta-compétences adaptées aux besoins futurs.
- **Globalisation** : émergence d'opportunités inédites à l'échelle mondiale [BN22].

1.3.4 Défis et perspectives

Bien que l'industrie 4.0 offre des perspectives de croissance économique grâce à l'amélioration de la productivité et à l'innovation rapide, elle présente également des défis majeurs :

- La cybersécurité.
- L'adaptation des compétences des travailleurs.
- L'investissement en infrastructures technologiques.
- La gouvernance des données et leur usage éthique.

1.4 Big Data et Cloud Computing au cœur de l'industrie 4.0

L'industrie 4.0 repose sur l'hyper-connectivité des systèmes, générant un important volume de données en temps réel, issues de sources multiples. Pour exploiter ces données, le recours à des technologies avancées est indispensable. Le cloud computing et le big data en constituent les piliers essentiels (Fig3.2), en fournissant les infrastructures et les capacités d'analyse nécessaires à la mise en œuvre de cette nouvelle ère industrielle [VEP18].

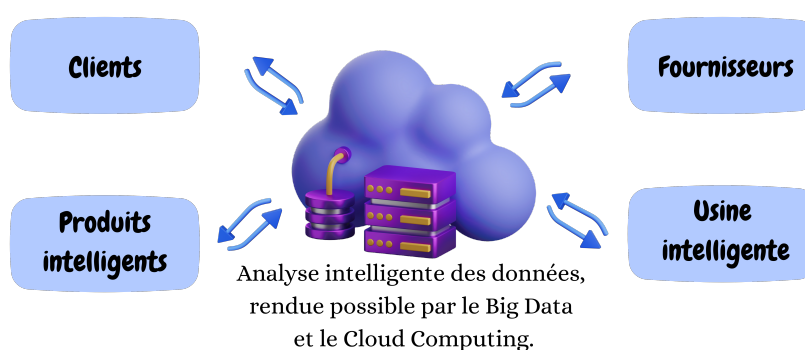


FIGURE 1.5 – Rôle du Big Data et du Cloud dans l'Industrie 4.0.

1.4.1 Cloud Computing

Le cloud computing est un pilier fondamental de l'industrie 4.0, permettant de stocker et traiter de grandes quantités de données générées par les machines et capteurs connectés. Il offre flexibilité et réduction des coûts en proposant des ressources informatiques à la demande, sans nécessiter d'infrastructure physique locale.

Ce modèle facilite l'accès aux données à tout moment et depuis n'importe où, encourageant la collaboration entre clients et fournisseurs, tout en impliquant davantage le client dans le processus de production.

Cependant, la latence due aux volumes importants de données peut nuire aux performances. Pour y remédier, le fog computing traite les données au plus près des objets connectés, ce qui réduit les délais, évite la congestion et améliore la qualité du service.

Des clouds privés peuvent être mis en place pour garantir la confidentialité, et des normes internationales (ISO/IEC) accompagnent les PME dans leur adoption de ces technologies [VEP18].

1.4.2 Big Data

Le Big data s'impose aujourd'hui comme un levier fondamental de la transformation numérique, en particulier dans le contexte de l'industrie 4.0. Apparue au début des années 2000 et popularisée à partir de 2011 grâce à son intégration au Gartner Hype Cycle, le concept s'est rapidement diffusé à travers des expressions telles que "les données sont le nouveau pétrole". Depuis, il a évolué d'une simple gestion de volumes massifs de données vers une discipline structurée autour des "5V" :

- **Volume** : la taille et l'échelle des données.
- **Variété** : la diversité des types et formats de données.
- **Vélocité** : la rapidité avec laquelle les données sont générées et doivent être traitées.
- **Véracité** : la qualité et la fiabilité des données.
- **Valeur** : l'utilité des données lorsqu'elles sont correctement contextualisées [Göll17, TN20].

Dans un environnement de production interconnecté, les machines, capteurs et systèmes intelligents génèrent en continu des flux de données à forte valeur ajoutée. Ces informations, lorsqu'elles sont correctement collectées, filtrées et analysées, permettent :

- D'optimiser les processus industriels en temps réel.
- D'améliorer la prise de décision à tous les niveaux hiérarchiques.
- D'anticiper les défaillances grâce à la maintenance prédictive.
- De réduire les coûts opérationnels et d'accroître la flexibilité face aux aléas

L'analyse avancée, descriptive, historique, prédictive ou prescriptive, ouvre la voie à une industrie agile, réactive et centrée client. Elle permet aussi la détection d'anomalies, l'adaptation aux conditions environnementales (humidité, température ...) ou encore l'amélioration des itinéraires logistiques.

Contrairement aux idées reçues, l'adoption du big data n'est pas liée à la taille de l'entreprise, mais à sa volonté stratégique d'exploiter les données pour améliorer ses performances. Cela

permet de concevoir de nouveaux produits ou services adaptés aux besoins réels des clients, de redéfinir les modèles économiques et de renforcer la compétitivité sur les marchés.

Malgré ses nombreux avantages, l'intégration du big data soulève plusieurs défis majeurs, en particulier :

- la qualité des données, indispensable à la fiabilité des analyses.
- la cybersécurité, avec des enjeux critiques liés à la confidentialité et à la protection des infrastructures.
- la gouvernance des données, qui devient un axe stratégique de pilotage [VEP18].

Conclusion

En conclusion, ce premier chapitre a posé les bases conceptuelles de l'industrie 4.0 en retraçant son évolution et en soulignant l'apport stratégique des technologies émergentes, notamment le big data. Il offre ainsi un cadre de référence essentiel pour aborder, dans les chapitres suivants, l'intégration concrète de ces technologies dans le cadre de notre étude de cas.

2

Data Science for Business

Introduction

Dans un environnement économique où les données occupent une place centrale, la capacité des entreprises à transformer ces données en leviers de performance devient un facteur clé de compétitivité. Ce chapitre s'inscrit dans la continuité du premier, qui a présenté les transformations liées à l'industrie 4.0 et les enjeux du Big Data. Il propose de s'intéresser plus spécifiquement aux apports de la data science dans le contexte des entreprises modernes.

L'objectif de ce chapitre est de présenter les fondements techniques et méthodologiques de la data science, en mettant en lumière ses principales composantes, les bases mathématiques, la visualisation des données, le ML, ainsi que la BI. Il mettra également en évidence la complémentarité entre BI et Data Science, et illustrera comment cette synergie peut contribuer à optimiser les processus métiers, notamment la chaîne d'approvisionnement.

Ce chapitre prépare ainsi le terrain pour la méthodologie data-driven, présentée dans le chapitre suivant, en fournissant les bases analytiques nécessaires à sa mise en œuvre concrète dans les organisations.

2.1 Data Science

La data science est une discipline interdisciplinaire qui combine les mathématiques, l'informatique, le ML et l'expertise métier. Elle permet de transformer les données en informations exploitables pour guider les décisions stratégiques. En analysant et en modélisant les données avec rigueur, tout en assurant leur interprétation à travers des visualisations claires, la data science contribue significativement à l'amélioration de la performance globale des organisations. Face à la croissance massive des données, elle s'impose aujourd'hui comme un outil clé pour les entreprises souhaitant gagner en efficacité, anticiper les tendances et optimiser leurs processus métiers.

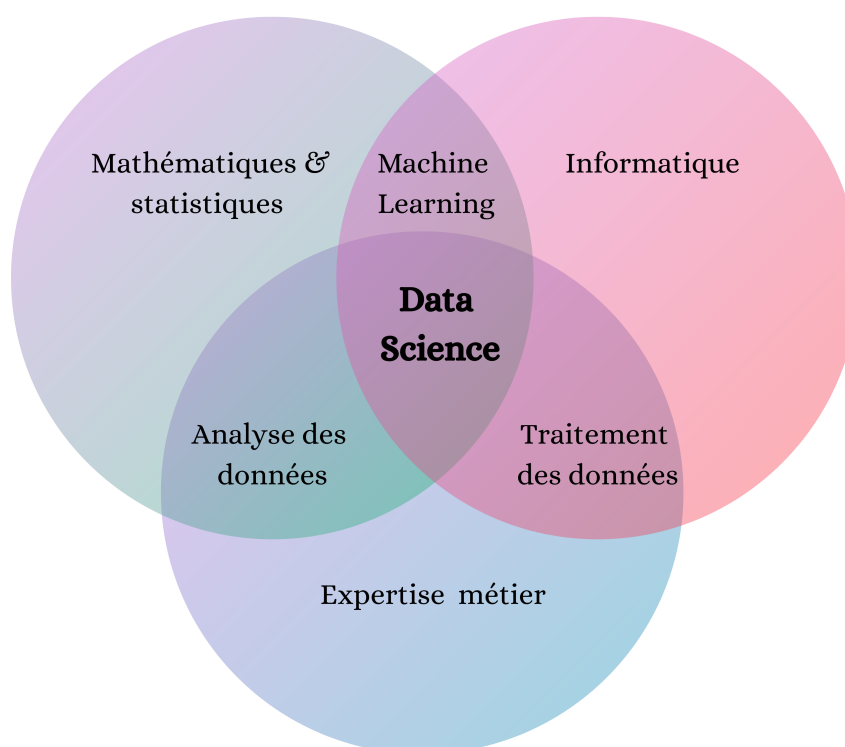


FIGURE 2.1 – Compétences et domaines clés en Data Science.

Le processus de la data science s’articule autour d’étapes fondamentales visant à convertir des données brutes en informations exploitables. Ces étapes comprennent généralement :

- **Définir le problème** : clarifier l’objectif à atteindre.
- **Collecter les données** : extraire les données pertinentes depuis différentes sources (bases de données, Interface de Programmation Applicative (API), web ...).
- **Préparer et explorer les données** : nettoyer, structurer et analyser pour identifier tendances et anomalies.
- **Modéliser et analyser** : concevoir et évaluer des modèles analytiques, qu’ils soient descriptifs, prédictifs ou prescriptifs, adaptés aux objectifs du projet, en mobilisant des outils statistiques, d’apprentissage automatique ou d’optimisation.
- **Communiquer les résultats** : utiliser des visualisations, rapports ou tableaux de bord pour appuyer la décision [Gru19, Ozd16].

2.1.1 Mathématiques pour la Data Science

La maîtrise des bases mathématiques est essentielle pour analyser, modéliser et interpréter les données de manière rigoureuse. Voici les principales disciplines mobilisées :

- **Algèbre linéaire** : utilisée pour représenter les données sous forme de vecteurs ou matrices, elle est essentielle dans le machine learning, notamment pour les systèmes de recommandation, la réduction de dimension et le traitement d’images.

- **Analyse mathématique** : elle permet d'optimiser les modèles grâce au calcul différentiel et intégral, notamment avec les dérivées utilisées dans les algorithmes de ML, comme la descente de gradient.
- **Probabilités et statistiques** : ces domaines servent à gérer l'incertitude, construire des modèles prédictifs et évaluer la fiabilité des résultats. Des notions avancées comme le théorème de Bayes, les intervalles de confiance ou les tests d'hypothèses sont régulièrement utilisées.
- **Géométrie analytique** : elle aide à mesurer les distances ou similarités entre données, notamment dans les algorithmes de classification, de clustering ou de détection d'anomalies.
- **Analyse numérique** : utilisée pour résoudre des équations complexes ou approchées, elle est essentielle dans les algorithmes nécessitant des calculs répétitifs ou itératifs.
- **Théorie des ensembles** : utile pour manipuler des groupes d'éléments distincts, elle intervient dans l'organisation des données et la logique de programmation des algorithmes [Ozd16].

2.1.2 Data Visualisation

La data visualisation est l'art de transformer des données complexes en représentations visuelles intuitives. Elle ne se limite pas à de simples graphiques, elle offre une perspective claire pour faciliter l'analyse et la prise de décision. En rendant les données plus compréhensibles, elle permet de mieux communiquer les résultats et d'en tirer des enseignements pertinents.

On distingue principalement deux usages de la data visualisation :

- **Explorer les données** : analyser visuellement les informations pour détecter des tendances, des corrélations ou des anomalies avant des analyses plus poussées.
- **Communiquer les résultats** : présenter les analyses de façon claire et accessible, adaptée au public cible.

Les principaux types de visualisations se répartissent en plusieurs catégories :

- **Graphiques de comparaison** : utilisés pour comparer différentes valeurs ou suivre leur évolution (diagramme en lignes, diagramme à barres, diagramme radar).
- **Graphiques de relation** : idéaux pour représenter les relations ou corrélations entre plusieurs variables (nuage de points, diagramme à bulles, corrélogramme, carte thermique).
- **Graphiques de composition** : permettent de visualiser les différentes parties d'un ensemble (diagramme circulaire, barres empilées, zone empilée, diagramme de Venn).
- **Graphiques de distribution** : illustrent la répartition des données et leurs variations (histogramme, courbe de densité, boîte à moustaches, diagramme en violon).
- **Graphiques géospatiaux** : utilisés pour représenter des données liées à des localisations géographiques (carte à points, carte choroplèthe, carte de connexions).
- **Tableaux de bord (Dashboards)** : outils interactifs combinant plusieurs visualisations sur un même écran. Ils offrent une vue synthétique des indicateurs clés et facilitent l'analyse, le suivi et la prise de décisions stratégiques [DG20].

2.2 Machine learning

Le ML, qui constitue une branche de l'IA, désigne la capacité des machines à apprendre automatiquement à partir des données, sans être explicitement programmées. Il représente une composante centrale de la Data Science, permettant de détecter des patterns, d'établir des prédictions ou d'expliquer des phénomènes complexes. Son objectif principal est d'exploiter les données disponibles pour anticiper des comportements futurs ou fournir des analyses décisionnelles pertinentes.

Le ML peut être classé selon plusieurs critères, tels que la nature des données utilisées, le degré d'intervention humaine nécessaire ou le mode d'apprentissage du modèle. Parmi les principales distinctions, on retrouve :

- **Selon la supervision des données :**
 - **Apprentissage supervisé** : basé sur des données étiquetées, il permet de réaliser des tâches de classification ou de régression (kNN, SVM, régressions, arbres de décision, réseaux de neurones...).
 - **Apprentissage non supervisé** : le modèle apprend sans données étiquetées. Il sert à détecter des structures cachées dans les données (k-means, PCA...).
 - **Apprentissage semi-supervisé** : combine un petit nombre de données étiquetées avec un grand volume non étiqueté (autoencodeurs, DBN...).
 - **Apprentissage par renforcement** : l'agent apprend par essais-erreurs en interagissant avec un environnement et en recevant des récompenses ou des pénalités (Policy Gradient, Actor-Critic...).
- **Selon la manière d'apprendre :**
 - **Apprentissage par lot (batch learning)** : le modèle est entraîné une seule fois sur l'ensemble des données disponibles. Toute mise à jour nécessite un nouvel entraînement complet.
 - **Apprentissage en ligne (online learning)** : le modèle est mis à jour progressivement au fil de l'arrivée de nouvelles données. Il est adapté aux flux continus et aux environnements à ressources limitées [Gru19, Gér17].

2.2.1 Démarche méthodologique d'un projet de ML

Le ML suit un processus méthodique pour convertir un problème métier en solution opérationnelle fondée sur les données. Les principales étapes sont :

Définition du problème et traitement des données

Afin de poser les bases d'un projet de ML, il est essentiel de définir clairement la problématique métier et les objectifs à atteindre. Cette première phase s'accompagne de la collecte des données pertinentes, de leur nettoyage et structuration, ainsi que d'une exploration préliminaire à l'aide de visualisations et de statistiques descriptives. L'objectif est d'assurer la qualité des données et de guider les choix techniques pour les étapes suivantes.

Préparation des données

Avant d'entraîner un modèle, il est nécessaire de préparer les données pour les rendre exploitables. Cette étape comprend plusieurs actions clés :

- **Transformation des variables catégorielles** : elle consiste à convertir les données qualitatives en format numérique. L'encodage des étiquettes attribue un chiffre à chaque catégorie, tandis que l'encodage One-Hot crée des colonnes binaires pour éviter toute hiérarchisation implicite des valeurs.
- **Création et sélection des variables** : la création et la sélection des variables permettent d'identifier les informations les plus pertinentes pour le modèle. Elles reposent sur l'expertise métier, des analyses statistiques et des tests. Des techniques comme la réduction de dimension ou la régularisation aident à simplifier l'apprentissage tout en évitant le manque ou l'excès de données.
- **Mise à l'échelle des caractéristiques** : la mise à l'échelle des variables est essentielle en machine learning, car de nombreux algorithmes sont sensibles aux écarts d'unité entre les données numériques. Deux techniques sont principalement utilisées :
 - **La normalisation** : qui ramène les valeurs dans un intervalle défini, généralement $[0, 1]$, est adaptée aux modèles comme les réseaux de neurones.
 - **La standardisation** : qui centre les données autour de zéro avec une variance unitaire, est généralement plus robuste face aux valeurs extrêmes.
- **Séparation des données** : elle consiste à distinguer les variables explicatives (features), qui servent d'entrée au modèle, de la variable cible (label), qui représente la sortie à prédire.

Choix et entraînement du modèle

Cette étape consiste à sélectionner les modèles adaptés, les entraîner, puis choisir le plus performant. Elle se déroule en plusieurs phases clés :

- **Sélection des modèles candidats** : choisir plusieurs algorithmes (simples et complexes) adaptés au type de tâche (classification, régression...).
- **Entraînement des modèles** : entraîner les modèles sur les données préparées. Ajuster les hyperparamètres si nécessaire pour améliorer les résultats.
- **Évaluation des performances** : mesurer la qualité des modèles à l'aide de métriques pertinentes et de validation croisée.
- **Gestion du surapprentissage** : contrôler l'overfitting en comparant les résultats sur les jeux d'entraînement et de validation, et appliquer des techniques comme la régularisation.
- **Comparaison et sélection finale** : comparer tous les modèles testés. Retenir celui qui offre le meilleur équilibre entre précision, robustesse et simplicité.
- **Sauvegarde du modèle** : conserver le modèle choisi pour son utilisation future et documenter les choix faits.

Optimisation et ajustement du modèle

Une fois le modèle choisi, l'objectif est d'en améliorer les performances grâce à l'ajustement de ses hyperparamètres. Plusieurs techniques sont utilisées :

- **Recherche d'hyperparamètres** : processus d'optimisation du modèle en ajustant ses paramètres. *Grid Search* teste toutes les combinaisons, tandis que *Randomized Search* en évalue un échantillon aléatoire pour plus de rapidité.
- **Méthodes par ensemble** : techniques qui consistent à combiner plusieurs modèles afin d'en former un plus performant. Cette approche permet souvent de surpasser les performances obtenues par un seul modèle pris isolément.
- **Analyse des erreurs et des variables** : étudier l'importance des variables et analyser les erreurs permet d'identifier les points faibles du modèle et d'améliorer sa précision.

Présentation des résultats

Il s'agit de présenter et documenter l'ensemble du travail :

- **Évaluation finale** : tester le modèle sur un jeu de données indépendant pour mesurer sa performance réelle.
- **Interprétation des résultats** : présenter les indicateurs clés (précision, taux d'erreur, RMSE...) de façon claire et accessible.
- **Synthèse et communication** : documenter la démarche et les résultats, et préparer des supports visuels pour faciliter la compréhension, y compris pour un public non technique.

Déploiement et maintenance du système

Une fois le modèle validé, sa mise en production nécessite un déploiement rigoureux et une maintenance continue afin d'assurer sa performance et sa fiabilité dans le temps :

- **Préparation au déploiement** : connexion aux sources réelles et mise en place de tests automatisés.
- **Surveillance et alertes** : suivi en temps réel des performances et détection des baisses de qualité.
- **Évaluation continue** : vérification régulière des prédictions par des experts et intégration des retours.
- **Contrôle de la qualité des données** : détection et correction des erreurs ou incohérences dans les données entrantes.
- **Mise à jour du modèle** : réentraînement périodique avec de nouvelles données et sauvegarde des versions fiables.
- **Maintenance technique** : procédures régulières pour assurer la stabilité, la fiabilité et la performance du système [Gru19, Gér17].

2.2.2 Métriques d'évaluation

L'évaluation d'un modèle de ML repose sur des métriques bien définies, sélectionnées selon la nature du problème à résoudre, classification ou régression.

Dans le cas de la classification, ces métriques s'appuient sur les notions de vrais positifs (TP), faux positifs (FP), vrais négatifs (TN) et faux négatifs (FN). Comme l'illustre le tableau ci-dessous (2.1), ces métriques permettent d'évaluer la qualité des prédictions [Dev24, Dat24] :

Métrique	Définition	Formule
Exactitude (Accuracy)	Proportion de bonnes prédictions parmi toutes les observations.	$\frac{TP + TN}{TP + TN + FP + FN}$
Précision (Precision)	Proportion de prédictions positives correctes parmi toutes les prédictions positives.	$\frac{TP}{TP + FP}$
Rappel (Recall)	Proportion des vrais positifs détectés parmi tous les positifs réels.	$\frac{TP}{TP + FN}$
F1-Score	Moyenne harmonique entre la précision et le rappel.	$2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$

TABLE 2.1 – Métriques d'évaluation pour les modèles de classification.

Pour la régression, les métriques mesurent l'écart entre les valeurs réelles observées (y_i) et les valeurs prédites (\hat{y}_i) sur un ensemble de n observations. Comme l'illustre le tableau ci-dessous (2.2), elles permettent d'évaluer la précision d'un modèle en comparant les prédictions aux valeurs observées [Dat24, Gér17] :

Métrique	Définition	Formule
Mean Absolute Error (MAE)	Moyenne des valeurs absolues des erreurs entre les valeurs réelles et prédites.	$\frac{1}{n} \sum_{i=1}^n y_i - \hat{y}_i $
Mean Squared Error (MSE)	Moyenne des carrés des erreurs, pénalisant davantage les grandes erreurs.	$\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$
Root Mean Squared Error (RMSE)	Racine carrée de la MSE, exprimée dans l'unité de la variable.	$\sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$
Coefficient de détermination (R^2)	Proportion de la variance des valeurs réelles expliquée par le modèle. Où \bar{y} est la moyenne des observations.	$1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$

TABLE 2.2 – Métriques d'évaluation pour les modèles de régression.

2.2.3 Défis majeurs du ML

Le développement d'un modèle performant en ML implique de surmonter plusieurs défis, notamment ceux liés à la qualité des données, à la complexité du modèle et à sa capacité de généralisation. Parmi les plus courants figurent :

— **Qualité et représentativité des données :**

- **Quantité insuffisante de données :** les algorithmes nécessitent un volume important d'exemples pour apprendre efficacement, notamment dans les cas complexes ou les modèles profonds.
- **Données non représentatives :** un échantillon biaisé ou mal équilibré empêche le modèle de généraliser correctement à de nouveaux cas.
- **Données de mauvaise qualité :** les erreurs, valeurs manquantes ou données bruitées perturbent l'apprentissage. Un prétraitement rigoureux est donc essentiel.
- **Caractéristiques peu pertinentes :** sans variables explicatives informatives, même un modèle sophistiqué échouera. Le feature engineering (création, transformation, ou sélection des variables) joue ici un rôle fondamental.

- **Surapprentissage (Overfitting)** : il survient lorsqu'un modèle apprend trop précisément les spécificités des données d'entraînement, y compris le bruit et les anomalies, ce qui nuit à sa capacité de généralisation. Plusieurs approches permettent de le prévenir :
 - Réduire la complexité du modèle (nombre de paramètres, profondeur).
 - Collecter plus de données pertinentes.
 - Nettoyer les données pour éliminer les erreurs et valeurs aberrantes.
 - Appliquer des techniques de régularisation (L1, L2, dropout. . .).
- **Sous-apprentissage (Underfitting)** : il désigne un modèle trop simple ou mal configuré, incapable de capturer les relations sous-jacentes dans les données. Il entraîne des performances faibles, même sur l'ensemble d'apprentissage. Plusieurs approches permettent de le prévenir :
 - Utiliser un modèle plus expressif (réseaux plus profonds, plus de neurones, . . .).
 - Introduire ou concevoir de meilleures variables (feature engineering).
 - Ajuster la régularisation pour améliorer la capacité d'apprentissage [Gér17].

2.3 Business Intelligence

La BI désigne un ensemble de technologies et de méthodes permettant de transformer les données en informations fiables, accessibles et compréhensibles pour faciliter la prise de décision. Elle vise à offrir aux utilisateurs des outils simples d'usage, sans nécessiter de compétences techniques, afin d'analyser les données selon des logiques métiers et soutenir les choix stratégiques des entreprises.

Plus qu'un simple ensemble d'outils, la BI s'inscrit dans une démarche globale reposant sur :

- **Des théories** : principes de conception et d'analyse.
- **Des méthodologies** : démarches structurées pour optimiser l'efficacité.
- **Une architecture** : organisation technique cohérente.
- **Des technologies** : logiciels, outils et infrastructures adaptés [NV17].

2.3.1 Méthodologie de la BI

La BI suit une méthodologie structurée visant à convertir les données brutes en informations décisionnelles, à travers un processus en plusieurs étapes :

1. **Collecte des données** : rassembler des données provenant de multiples sources internes et externes.
2. **Intégration des données** : elle repose sur un processus clé (Extract Transform Load (ETL)) qui comprend trois phases fondamentales :
 - **Extraction** : récupération des données brutes collectées.
 - **Transformation** : nettoyage, formatage et enrichissement des données pour assurer leur qualité et leur cohérence.
 - **Chargement** : insertion des données transformées dans un entrepôt de données centralisé (Datawarehouse (DWH)).

3. **Stockage et modélisation** : une fois centralisées, les données sont organisées dans le Datawarehouse selon des modèles adaptés (étoile, flocon, constellation), afin de faciliter l'analyse et d'assurer de bonnes performances.
4. **Analyse des données** : exploiter les données à travers des analyses descriptives ou exploratoires, en s'appuyant sur des outils tels que l'Online Analytical Processing (OLAP), le Data Mining et des méthodes statistiques, afin d'en extraire des insights pertinents pour la décision.
5. **Restitution et visualisation** : enfin, les résultats des analyses sont présentés sous forme de tableaux de bord, rapports interactifs ou graphiques.
6. **Partage des résultats** : les tableaux de bord et rapports générés sont ensuite partagés avec les parties prenantes à travers des outils collaboratifs. Ce partage facilite la communication, renforce l'alignement stratégique et favorise une prise de décision collective et éclairée [NV17].

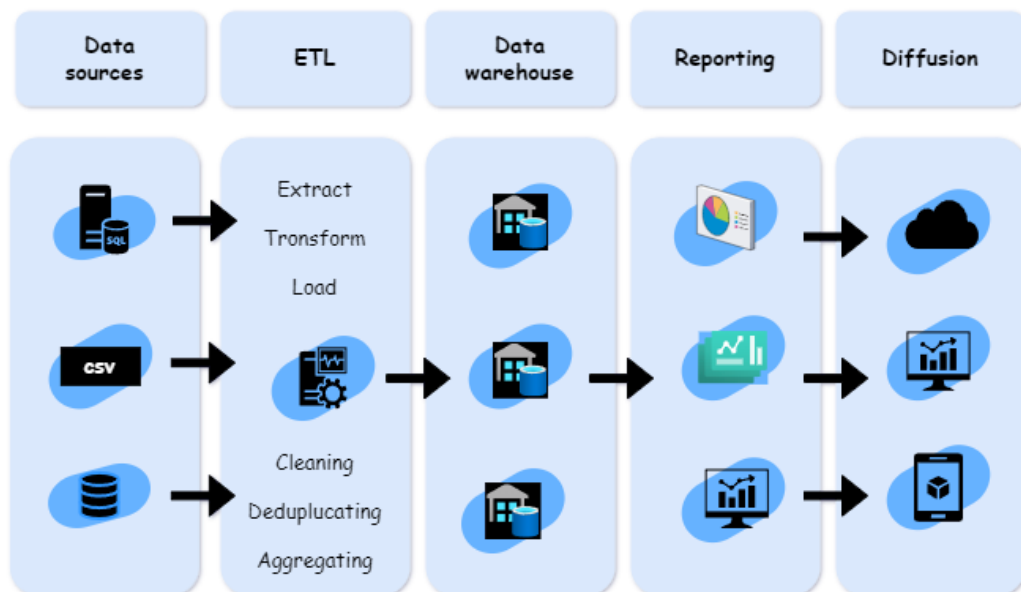


FIGURE 2.2 – Cycle de vie des données.

2.3.2 Architecture de stockage

La BI repose sur une architecture structurée combinant un entrepôt central (Datawarehouse) et des sous-ensembles spécialisés (DataMarts) dédiés, afin de garantir une analyse fiable, cohérente et adaptée aux besoins métiers :

- **DWH** : est l'entrepôt central de la BI, conçu pour centraliser, structurer et fiabiliser les données issues de sources multiples. Il permet de conserver un historique, d'analyser les tendances et de produire des prévisions, grâce à une organisation rigoureuse fondée sur

des règles métiers et des types de tables spécifiques. Son actualisation peut être périodique ou quasi temps réel selon les besoins.

- **DataMart** : est un sous-ensemble thématique du DWH, conçu pour répondre aux besoins spécifiques d'un domaine métier (ventes, finance...). Il offre un accès ciblé et rapide aux données pertinentes, facilitant ainsi les analyses opérationnelles. Bien qu'il repose sur les mêmes principes qu'un Datawarehouse, sa portée est plus restreinte et orientée vers un usage métier spécifique [NV17].

2.3.3 Modélisation des données en BI

La modélisation des données constitue une étape clé dans la mise en œuvre d'une solution de BI. Elle vise à structurer les informations de manière cohérente et optimisée afin de répondre aux exigences d'analyse et de pilotage, en s'appuyant sur trois niveaux complémentaires :

1. **Modèle logique** : il constitue la première étape dans la conception d'un DWH. Il définit les entités métier (clients, produits, ventes...) et leurs relations, assurant une structuration cohérente avec les objectifs d'analyse. Véritable lien entre les besoins métiers et les contraintes techniques, il garantit l'exhaustivité et la clarté des données à exploiter. Il s'appuie généralement sur trois types de tables principales :
 - **Tables de faits** : regroupent les données quantitatives à analyser (ventes, coûts, marges...), parfois agrégées pour améliorer les performances.
 - **Tables de dimensions** : fournissent les informations descriptives associées aux faits (produits, clients, périodes, lieux...).
 - **Tables de relations** : permettent de gérer les liens entre différentes entités (calendriers, hiérarchies...).
2. **Modèle relationnel** : il structure les données sous forme de tables liées entre elles par des clés primaires et étrangères. C'est le schéma le plus courant dans les bases de données relationnelles, notamment en BI. Plusieurs variantes sont utilisées selon les objectifs :
 - **Modèle normalisé** : minimise la redondance, adapté aux environnements transactionnels.
 - **Modèle dénormalisé** : introduit des duplications contrôlées pour améliorer les performances d'analyse.
 - **Modèle en étoile** : table de faits centrale connectée à des dimensions, modèle de référence pour les DataMarts.
 - **Modèle en flocon** : version normalisée du modèle en étoile, mieux adapté aux structures hiérarchiques complexes.
 - **Modèle en constellation** : intègre plusieurs tables de faits partageant des dimensions communes, facilitant l'analyse croisée de plusieurs processus métiers.
3. **Modèle physique** : Le modèle physique est la concrétisation technique du modèle logique. Il définit précisément la manière dont les données sont stockées dans la base, en tenant compte des contraintes de performance et d'intégrité. Il inclut :
 - La définition des champs et de leurs types de données.
 - L'implémentation des clés primaires et étrangères.
 - La gestion de l'intégrité référentielle.
 - L'optimisation des traitements ETL et des requêtes analytiques [NV17].

2.4 Synergie entre BI et Data Science

Pendant longtemps, la BI a été au cœur des systèmes d'aide à la décision, des premiers outils des années 1970 jusqu'aux solutions modernes en self-service comme Power BI, Tableau ou Excel OLAP. Cependant, avec l'explosion des volumes de données et la complexité croissante des environnements numériques, les limites des approches BI traditionnelles sont devenues évidentes.

Aujourd'hui, l'intégration croissante de l'IA dans la BI représente une réponse stratégique à plusieurs enjeux :

- **Réactivité accrue** : les utilisateurs attendent des réponses immédiates, formulées simplement, comme avec Alexa ou Google. L'IA permet de générer automatiquement des analyses compréhensibles en langage naturel.
- **Accès simplifié à l'analyse** : face à la masse croissante de données, les outils classiques deviennent trop complexes pour les non-spécialistes. L'IA rend possible l'analyse avancée même sans compétences techniques.
- **Généralisation du ML** : les plateformes low-code ou no-code rendent les algorithmes de ML accessibles à tous, poussant les équipes BI à intégrer des prévisions et recommandations dans leurs tableaux de bord.

D'après le modèle proposé par Gartner, l'analyse de données s'organise en quatre niveaux de maturité, descriptif, diagnostique, prédictif et prescriptif. La BI maîtrise les deux premiers niveaux, essentiels pour comprendre le passé. La Data Science, quant à elle, intervient sur les niveaux prédictif et prescriptif, apportant des capacités d'anticipation et d'aide à la décision.

Pourtant, dans de nombreuses entreprises, les équipes BI et Data Science restent cloisonnées, avec des outils, des langages et des objectifs différents. Là où le data scientist parle de modèles et d'algorithmes, l'expert BI évoque pipelines ETL et cubes OLAP. Cette séparation freine la synergie, alors qu'une meilleure collaboration permettrait d'accélérer la transformation data-driven.

En réalité, la BI et la Data Science sont complémentaires. La BI offre une vision claire et structurée du passé, tandis que la Data Science ouvre la voie à l'anticipation et à l'optimisation. Ensemble, elles permettent de dépasser le simple reporting pour construire une intelligence décisionnelle augmentée [Zwi22].

2.5 Optimisation de la chaîne d'approvisionnement

La planification de la chaîne d'approvisionnement joue un rôle stratégique dans la maîtrise des coûts, le financement de la croissance et l'amélioration de la compétitivité. Elle repose sur une coordination étroite entre la finance et la logistique, visant à optimiser les flux d'informations, de biens et financiers. Des outils comme la cartographie des flux, la modélisation et le calcul du coût total d'acquisition permettent d'atteindre un équilibre entre rentabilité, performance opérationnelle et gestion de la trésorerie, tout en intégrant les contraintes externes.

2.5.1 Processus d'optimisation de la chaîne d'approvisionnement

L'optimisation de la chaîne d'approvisionnement repose sur une planification structurée visant à aligner la demande, la production et les approvisionnements. Elle permet d'assurer la disponibilité des produits tout en maîtrisant les coûts et les délais.

Planification de la demande

Première étape clé de l'optimisation de la chaîne d'approvisionnement, la planification de la demande vise à anticiper les besoins futurs des clients afin d'ajuster efficacement les approvisionnements. Grâce à l'utilisation de méthodes avancées telles que les séries temporelles et les algorithmes de ML, les prévisions gagnent en précision et en réactivité face aux évolutions du marché.

Planification de la production

La planification de la production vise à garantir la disponibilité des produits en conciliant les prévisions de la demande, les capacités de production et les ressources disponibles. Elle repose sur une organisation structurée permettant d'éviter les ruptures, d'optimiser les coûts et d'assurer la continuité des opérations.

Au cœur de cette démarche se trouve le Plan Industriel et Commercial (PIC), qui offre une vision globale et anticipée des besoins, assure la coordination entre les services concernés et aligne les opérations avec la stratégie de l'entreprise.

Sur la base du PIC, le Plan Directeur de Production (PDP) traduit ces prévisions en programmes opérationnels précis, il définit les produits à fabriquer, les quantités nécessaires et les périodes de production en tenant compte des capacités réelles.

Cette planification s'appuie sur des indicateurs de performance, un suivi rigoureux des écarts et une gestion proactive des stocks. Elle permet une vision cohérente à moyen et long terme, facilite la coordination interfonctionnelle et soutient les prises de décision éclairées.

La faisabilité du plan repose sur la prise en compte de plusieurs contraintes clés :

- Disponibilité des ressources humaines (heures de travail).
- Capacité réelle des machines (maintenance, pannes).
- Contraintes budgétaires et logistiques.
- Exigences environnementales et de sécurité.

La planification de la production s'adapte aux contextes industriels selon trois principales approches :

- **Fabrication sur stock** : production anticipée pour constituer un stock de sécurité.
- **Fabrication à la commande** : production déclenchée uniquement à la réception d'une commande.
- **Assemblage à la commande** : assemblage final réalisé à partir de composants déjà disponibles.

L'optimisation de la production repose sur plusieurs méthodes :

- **Théorie des Contraintes (TOC)** : identifie et gère les points de blocage pour améliorer le flux global.
- **Modèle de Wilson (EOQ)** : calcule la quantité optimale à commander en équilibrant coûts de stockage et d'approvisionnement.
- **Modèles d'optimisation mathématique** : résolvent des problèmes complexes de planification et de répartition des ressources selon des objectifs prédéfinis.

Planification des approvisionnements et gestion des stocks

La planification des approvisionnements vise à garantir la disponibilité des MP au bon moment, en quantité suffisante et à un coût maîtrisé, assurant ainsi la continuité de la production et la satisfaction des clients.

- **MRP** : il permet d'anticiper les besoins en MP à partir du PDP, des nomenclatures produits et des stocks disponibles. Il facilite la planification des commandes fournisseurs, des ordres de fabrication et la génération automatique des documents nécessaires. Ses principaux avantages sont les suivants :
 - Réduction des stocks excédentaires.
 - Limitation des ruptures et arrêts de production.
 - Optimisation des coûts d'achat et de stockage.
 - Amélioration de la réactivité face aux variations de la demande.
- **Rôle des stocks** : ils assurent l'équilibre entre l'offre et la demande. Leur gestion vise à optimiser la disponibilité tout en limitant les coûts liés à l'entreposage, au financement et aux risques de dépréciation ou d'obsolescence.
- **Stratégies d'approvisionnement** :
 - **Approvisionnement push** : basé sur les prévisions et planifié par le MRP.
 - **Approvisionnement pull** : déclenché par la consommation réelle.
 - **VMI** : le fournisseur gère le réapprovisionnement en fonction des besoins du client.
- **Outils d'optimisation** :
 - Modèle de Wilson (Quantité Économique à Commander (EOQ)).
 - Les systèmes de réapprovisionnement à seuil (point de commande).

La planification des approvisionnements et des stocks vise à garantir une disponibilité optimale tout en équilibrant performance de service et coûts globaux [For19].

Conclusion

Ce chapitre a permis d'explorer le rôle stratégique de la data science dans l'univers de l'entreprise, en détaillant ses fondements, ses outils et ses applications concrètes. En articulant analyse statistique, visualisation et le ML, la data science se positionne comme un levier essentiel pour enrichir la compréhension des données et anticiper les dynamiques métier. Elle complète ainsi la BI, en élargissant le spectre de l'analyse vers des dimensions prédictives et prescriptives.

L'approche intégrée présentée dans ce chapitre montre que la convergence entre BI et data science est un facteur clé d'accélération pour toute entreprise souhaitant s'engager dans une démarche data-driven. Ces éléments conceptuels et méthodologiques poseront les fondations de la suite de ce mémoire, qui proposera une méthodologie de transformation organisationnelle, avant d'illustrer son application concrète à travers l'étude de cas menée chez Ovitalé.

3

Méthodologie Data-Driven

Introduction

Dans un contexte industriel en pleine mutation, où les données deviennent un levier stratégique fondamental, la transition vers une organisation data-driven s'impose comme une priorité pour les entreprises souhaitant rester compétitives. Ce chapitre s'inscrit dans la continuité des deux premiers volets de ce mémoire, après avoir exploré les fondements de l'industrie 4.0 et les bases théoriques de la data science et la BI, il s'agit ici de présenter concrètement la méthodologie permettant de mettre en œuvre une telle transformation organisationnelle.

L'objectif principal de ce chapitre est de décrire les différentes étapes clés pour adopter une approche pilotée par les données, en s'appuyant sur des concepts, des outils et des pratiques éprouvées. Il vise également à fournir une vision opérationnelle des conditions de réussite de cette démarche, en mettant en lumière les dimensions techniques, humaines et organisationnelles.

En ce sens, ce chapitre prépare le terrain pour le chapitre suivant, dédié à notre étude de cas, où la méthodologie présentée ici sera mise en œuvre dans un contexte réel.

3.1 Le concept de Data-Driven

La notion de data driven (piloté par les données) est largement abordée à la fois dans la littérature professionnelle et la littérature organisationnelle. Les termes "data driven" et "organisation pilotée par les données" sont utilisés de manière interchangeable [BM24].

3.1.1 Dans la littérature professionnelle

Les rapports de cabinets de conseil en stratégie, organisation, technologie et management, comme McKinsey & Company, Boston Consulting Group (BCG), Deloitte ou encore Capgemini, emploient fréquemment le terme "data driven" sans en fournir une définition explicite. Ils introduisent également de nouveaux concepts tels que "data-powered enterprise" ou "insight-driven organization", avec un objectif commun, intégrer les données dans la culture d'entreprise, les processus décisionnels et les démarches d'innovation.

Un point essentiel souligné par ces rapports est l'importance de la culture d'entreprise. Une enquête menée par le cabinet de conseil américain spécialisé dans la stratégie de données et l'IA, NewVantage Partners, révèle que les principaux freins à la transformation data-driven sont d'ordre culturel et humain, plutôt que technologique. Une transformation réussie repose ainsi sur un leadership fort et des initiatives internes cohérentes.

La prise de décision basée sur les données est centrale, bien que les publications divergent sur l'équilibre entre données et intuition. Certaines analyses insistent sur la nécessité de combiner rationalité analytique et expérience humaine [BM24].

3.1.2 Dans la littérature organisationnelle

Une étude menée par Brynjolfsson en 2011 a examiné 179 grandes entreprises cotées en bourse et a révélé que celles qui mettaient en œuvre une prise de décision basée sur les données étaient de 5 à 6 % plus productives que leurs concurrents.

Plusieurs chercheurs ont proposé des cadres conceptuels pour définir ce qu'est une organisation data-driven :

- Collecte de données, compétences, culture, prise de décision basée sur les données.
- Collecte de données, capacités analytiques, culture data-driven, prise de décision et création de valeur.
- La culture analytique et l'intégration des analyses prédictives et prescriptives.

Ces approches s'appuient sur les trois types d'analytique :

- **Analytique descriptive** : ce qui s'est passé (rapports, visualisations, ...).
- **Analytique prédictive** : ce qui va se passer (modèles prédictifs, ML, ...).
- **Analytique prescriptive** : ce qu'il faut faire (optimisation, simulations, ...).

Les deux derniers types (prédictif et prescriptif) sont souvent qualifiés d'analyse avancée et sont essentiels pour qu'une organisation soit véritablement data-driven [BM24].

3.2 Définition d'une organisation data-driven

Tout d'abord, avant de définir ce qu'est une organisation data-driven, il convient de comprendre ce qu'est une organisation non data-driven.

Une organisation non data-driven base principalement ses décisions sur l'intuition et l'expérience personnelle uniquement, sans s'appuyer systématiquement sur des analyses de données.

Elle peut également se référer à des habitudes historiques ou à des structures hiérarchiques traditionnelles. Les données peuvent exister dans ce type d'organisation, mais elles sont souvent sous-utilisées, cloisonnées dans des silos, ou mobilisées de manière ponctuelle sans analyse approfondie. Ce type d'organisation se caractérise par un manque d'outils analytiques, une faible culture des données et une prise de décision plus réactive que proactive. En conséquence, elle peut rencontrer des difficultés à anticiper les évolutions du marché, à innover rapidement ou à optimiser ses processus internes de manière structurée.

Une organisation data-driven est une entité qui fonde ses décisions, ses actions et ses stratégies sur l'analyse rigoureuse de données plutôt que sur l'intuition ou l'expérience seule. Elle se caractérise par une culture organisationnelle dans laquelle tous les acteurs adhèrent à l'usage des données comme fondement principal à la prise de décision.

Elle s'appuie sur :

- Une gouvernance solide des données.
- Des technologies adaptées favorisant l'accès en libre-service aux informations.
- Des compétences analytiques réparties dans toute l'organisation.
- Une culture de l'expérimentation où les données priment sur les opinions, et où l'échec est accepté tant qu'il permet un apprentissage.

L'organisation data-driven exploite les données de manière opportune pour :

- Optimiser ses processus internes.
- Développer des produits ou services innovants.
- Accroître son avantage concurrentiel.
- Orienter ses décisions stratégiques par des modèles prédictifs et prescriptifs.

Être data-driven, c'est être une organisation pilotée par les données, capable de transformer celles-ci en valeur, en actions concrètes, et en facteurs de croissance durable.

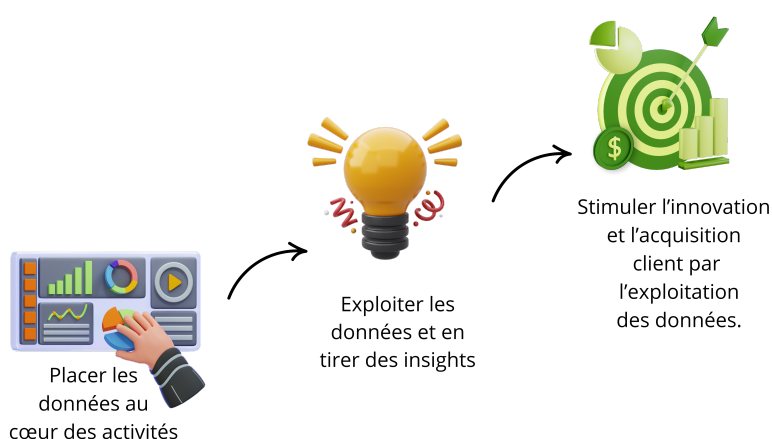


FIGURE 3.1 – Organisation data-driven.

3.3 Les fondations d'une organisation data-driven

Le cabinet *Projexion*, spécialisé en conseil en organisation et en transformation digitale, souligne que pour devenir une organisation data-driven, il est essentiel de mobiliser trois dimensions imbriquées :

- **Organisation** : il s'agit d'adapter la structure et la gouvernance pour intégrer les données dans tous les processus métiers. Cela implique de casser les silos, responsabiliser les équipes, et favoriser la collaboration entre les fonctions.
- **Données** : la data devient un actif stratégique. Elle doit être placée au cœur des priorités métiers, avec une vision claire, des compétences en data literacy, et des usages ancrés dans les pratiques quotidiennes.
- **Technologie** : un système d'information aligné avec les enjeux data et métiers est indispensable. Il doit permettre de collecter, traiter et exploiter les données de manière fluide, sécurisée et évolutive [Pro25].



FIGURE 3.2 – Les fondations d'une organisation data-driven.

3.4 Processus de transformation vers une organisation data-driven

3.4.1 Analyse des besoins

L'analyse des besoins est la première étape essentielle dans la mise en place d'une stratégie data-driven. Elle permet d'identifier les écarts entre les performances actuelles et les objectifs à atteindre, en repérant les lacunes ou opportunités d'amélioration dans les processus existants. Cette analyse aide à comprendre quels sont les besoins spécifiques en termes de données et d'outils nécessaires pour optimiser les processus métiers.

Une fois les besoins identifiés, il est crucial de définir des objectifs spécifiques et mesurables. Ces objectifs doivent être alignés avec la stratégie globale de l'organisation. L'analyse des besoins aide également à déterminer les KPIs qui permettront de suivre les progrès et de valider l'efficacité des actions mises en place [Act25].

3.4.2 Collecte des données

Dans une démarche data-driven, la qualité et la diversité des données collectées sont essentielles pour alimenter les analyses, les modèles prédictifs et soutenir la prise de décision. Il ne s'agit pas seulement de collecter ce que l'on possède déjà, mais aussi d'explorer des sources externes, gratuites ou payantes, qui peuvent enrichir la compréhension des comportements et des contextes métier. Le processus de collecte des données doit donc être stratégique, structuré et orienté vers la valeur que ces données peuvent générer.

Les organisations peuvent mobiliser plusieurs types de données pour optimiser leurs analyses :

- **Données internes** : provenant des systèmes internes (ventes, stocks, clients...).
- **Données contextuelles** : permettant d'expliquer ou prédire certains comportements (météo, jours fériés...).
- **Données publiques** : accessibles librement via des portails gouvernementaux ou open data.
- **Données achetées** : fournies par des tiers spécialisés (Datalogix, Axciom, Experian...), parfois exclusives ou de meilleure qualité que les données internes.

Il est recommandé de collecter autant de données que possible dès le départ, même si leur utilité immédiate n'est pas évidente, car cela permet de fournir du contexte à certaines actions (comme une commande influencée par un événement extérieur) et d'avoir une vision plus complète des comportements des clients et des processus métiers [And15].

Avant d'acheter des données, il est important de considérer plusieurs critères :

- **Prix** : est-ce un bon investissement ?
- **Qualité** : les données sont-elles propres et fiables ?
- **Exclusivité** : sont-elles uniques ou accessibles à tous ?
- **Échantillon** : peut-on tester un échantillon avant d'acheter ?
- **Mises à jour** : à quelle fréquence sont-elles actualisées ?
- **Fiabilité technique** : API stable, disponibilité, limitations.
- **Sécurité** : transmission chiffrée, conformité aux normes.
- **Conditions d'utilisation** : y a-t-il des restrictions juridiques ?
- **Format** : CSV, JSON, XML sont préférables pour la flexibilité.
- **Documentation** : une bonne documentation avec dictionnaire de données est essentielle.
- **Volume et granularité** : correspondent-ils aux besoins métiers ?

3.4.3 Qualité des données

Les données sont souvent imparfaites, avec des biais cachés qui peuvent fausser les analyses, et leur traitement (nettoyage, structuration, préparation) s'avère souvent complexe, long et coûteux. C'est pourquoi les data scientists consacrent généralement 80 % de leur temps à préparer les données, et seulement 20 % à les analyser ou à construire des modèles.

Même avec un grand volume de données, une organisation n'est pas nécessairement pilotée par les données. Contrairement aux promesses marketing de certains fournisseurs de big data, la quantité ne garantit pas la valeur, une petite quantité de données propres et fiables est bien plus précieuse que des pétaoctets de données inutilisables.

La qualité des données ne peut être évaluée par un seul indicateur, car elle repose sur plusieurs dimensions :

- **Accessibilité** : les données doivent être facilement consultables par les analystes, via les bons outils et avec les autorisations nécessaires.
- **Exactitude** : les données doivent refléter fidèlement la réalité (ex : adresses à jour, mesures correctes ...).
- **Cohérence** : les données doivent pouvoir être liées entre elles de manière logique (ex : une commande liée à un client, une adresse ...).
- **Complétude** : les données doivent être entièrement présentes, sans valeurs ou enregistrements manquants.
- **Consistance** : les données doivent être uniformes entre différentes sources (ex : même email pour un client dans tous les systèmes).
- **Définition claire** : chaque champ doit avoir une signification explicite et bien documentée.
- **Pertinence** : les données doivent être utiles et liées à l'analyse visée.
- **Fiabilité** : les données doivent être à la fois exactes et complètes.
- **Actualité** : les données doivent être disponibles dans un délai raisonnable pour permettre leur exploitation.

Un défaut dans l'une de ces dimensions peut rendre les données inexploitable, incomplètes ou trompeuses. La gestion de la qualité des données est donc une responsabilité partagée entre tous les acteurs de la chaîne analytique [And15].

3.4.4 Accès aux données

Pour être véritablement pilotée par les données, une organisation ne doit pas seulement disposer de données précises, pertinentes et à jour. Elles doivent également être :

- **Joinable** : pour une analyse efficace, les données doivent être associables à d'autres sources via des outils adaptés comme les bases de données relationnelles, NoSQL ou Hadoop. L'utilisation d'Excel devient vite limitée avec de gros volumes de données, ralentissant les calculs et la productivité. En centralisant les données dans une base comme MySQL et en automatisant les requêtes, les analystes gagnent en efficacité et en qualité d'analyse.

- **Partageables** : pour qu'une organisation soit réellement pilotée par les données, il est essentiel d'adopter une culture de partage. Les données doivent être partageables afin de permettre leur croisement entre différents services ou systèmes. Par exemple, dans une entreprise industrielle, si les données de production ne sont pas partagées avec les équipes de maintenance et de logistique, il devient difficile de prévenir les pannes, d'optimiser les stocks ou de planifier les livraisons efficacement. Le cloisonnement des données freine ainsi l'amélioration des processus, tandis que leur mutualisation permet une vision globale, bénéfique à l'ensemble de la chaîne de valeur.
- **Interrogeables** : il est essentiel de disposer d'outils adaptés permettant d'interroger, de filtrer, de trier et d'agréger les données. Le reporting et l'analyse reposent sur la capacité à transformer de grandes quantités de données brutes en indicateurs synthétiques et pertinents, afin de mieux comprendre les dynamiques de l'entreprise. Il doit être possible d'identifier des tendances, d'analyser les écarts entre différents segments de clients, ou encore d'explorer des axes d'amélioration. Pour cela, les analystes doivent avoir accès à des outils leur facilitant la génération de ces indicateurs de manière rapide et efficace.

Avoir des données accessibles ne suffit pas, des experts doivent savoir les exploiter efficacement. Les décisions doivent être guidées par des indicateurs bien définis et interprétés par l'humain [And15].

3.4.5 Organisation des équipes analytiques

La structuration d'une organisation orientée vers les données repose non seulement sur les technologies déployées, mais surtout sur la constitution d'équipes analytiques solides, multidisciplinaires et complémentaires. À ce titre, on peut en citer les trois principaux grands profils, chacun jouant un rôle clé dans la chaîne de valeur des données :

- **Le data engineer** : chargé de la conception, de la gestion et de l'optimisation des infrastructures de données, assurant la fiabilité, la scalabilité et l'accessibilité des flux de données.
- **Le data analyst** : qui extrait, nettoie et analyse les données pour produire des rapports, des tableaux de bord et des indicateurs facilitant la prise de décision.
- **Le data scientist** : qui mobilise des techniques avancées de machine learning et de modélisation statistique pour générer des prédictions, automatiser des processus et découvrir des patterns complexes.

Les métiers de la data forment un domaine en perpétuelle évolution, marqué notamment, en 2025, par l'émergence de nouveaux profils tels que le data ethicist, garant de l'éthique et de la responsabilité dans l'usage des données ou encore le data storyteller, dont le rôle est de transformer les analyses complexes en récits clairs et accessibles pour les décideurs. Cette dynamique illustre l'importance stratégique croissante des données au sein des organisations, et impose une adaptation continue des compétences et des fonctions au sein des équipes analytiques.

La structuration de ces rôles varie selon la taille et la complexité des organisations, dans les grandes entreprises, ils tendent à être bien définis et spécialisés, tandis que dans des structures plus petites, ces missions sont souvent regroupées sous des profils polyvalents aux responsabilités élargies [And15].

3.4.6 Du reporting à l'analyse

La transition du reporting vers l'analyse reflète une évolution stratégique, il ne s'agit plus seulement de constater les faits, mais d'en tirer des insights pour anticiper et orienter les décisions de manière proactive.

- **Reporting** : il s'agit de la présentation structurée de données passées, souvent sous forme de tableaux de bord ou de rapports, pour suivre la performance d'une entreprise. Il répond à la question « Que s'est-il passé ? ».
- **Alerting (Alertes)** : ce sont des notifications en temps réel signalant une anomalie ou un événement inhabituel. Elles répondent à « Que se passe-t-il maintenant ? », mais sans expliquer pourquoi.
- **Analysis (Analyse)** : c'est le processus qui consiste à comprendre les causes derrière les données observées, à interpréter les tendances et à proposer des actions ou prédictions. Elle répond aux questions « Pourquoi ?, Que faire ? ou Que va-t-il se passer ? ».

Le reporting et les alertes sont essentiels mais insuffisants dans une approche data-driven, car ils restent descriptifs et sans contexte. L'analyse est la véritable clé, elle donne du sens aux données, explique les évolutions, identifie les leviers d'action, et permet de formuler des décisions stratégiques basées sur des modèles ou expérimentations. Pour qu'une entreprise soit réellement pilotée par les données, elle doit aller au-delà de l'observation vers une compréhension approfondie et une capacité à anticiper et optimiser [And15].

3.4.7 Analyse de données

Les solutions d'analyse de données offrent aux entreprises un levier essentiel pour valoriser leurs données. Cependant, la grande diversité d'outils disponibles sur le marché peut semer la confusion, car beaucoup relèvent de catégories d'analyse différentes. Alors, comment s'y retrouver ? Voici les quatre principales formes d'analyse de données :

- **Analyse descriptive** : elle examine les données historiques afin de fournir une vue d'ensemble de ce qui s'est passé. Elle permet de mieux comprendre les performances passées, en contextualisant les événements à travers des graphiques, tableaux de bord ou rapports. Cette analyse répond à la question « Que s'est-il passé ? » et constitue la base de toute démarche analytique.
- **Analyse diagnostique** : elle cherche à comprendre les raisons derrière les événements observés. En répondant à la question « Pourquoi cela s'est-il produit ? », elle fait appel à des techniques telles que l'exploration de données (data mining), les analyses détaillées (drill-down, drill-through) et la recherche de corrélations. Elle permet d'identifier les causes racines pour guider les décisions futures.
- **Analyse prédictive** : elle utilise les données historiques pour anticiper les résultats futurs. Grâce à des modèles de ML, à l'IA et à des méthodes statistiques comme la régression, cette analyse répond à la question « Que va-t-il se passer ? » et aide les entreprises à planifier plus efficacement.
- **Analyse prescriptive** : elle va au-delà de la prédiction en recommandant les actions à entreprendre. Elle répond à la question « Que devons-nous faire ? ». C'est la forme la plus avancée d'analyse, reposant sur des algorithmes d'optimisation et de simulation. Elle est

notamment utilisée pour améliorer la tarification, la gestion des stocks ou la planification des ressources.

Les entreprises les plus avancées combinent ces types d'analyses complémentaires afin de prendre des décisions éclairées, basées sur des données fiables et contextualisées.

Par ailleurs, un nouveau type d'analyse émerge :

- **L'analyse augmentée** : elle combine l'IA et le ML pour automatiser certaines étapes de l'analyse, comme le nettoyage ou la préparation des données. Elle suggère des visualisations pertinentes, détecte automatiquement des tendances ou anomalies, et démocratise l'usage de la data grâce à des interfaces intuitives. Elle est particulièrement utile dans les environnements de Big Data, où les volumes et la complexité des données nécessitent une approche assistée.

Pour mieux comprendre l'utilité de chaque type d'analyse, le tableau ci-dessous (Tab 3.1) propose une synthèse accompagnée d'exemples concrets d'application :

Type d'analyse	Question clé	Exemple
Analyse descriptive	Que s'est-il passé ?	Analyse des ventes par catégorie de produits et comparaison avec les périodes précédentes.
Analyse diagnostique	Pourquoi cela s'est-il produit ?	Expliquer pourquoi certains produits se sont mieux vendus que d'autres.
Analyse prédictive	Que va-t-il se passer ?	Prédire une panne de machine grâce à un modèle basé sur les données historiques de maintenance.
Analyse prescriptive	Que devons-nous faire ?	Recommander la meilleure façon d'utiliser le budget publicitaire pour booster les ventes.

TABLE 3.1 – Les quatre principaux types d'analyse de données

Une stratégie d'analyse efficace repose sur une évaluation régulière des résultats obtenus. Cette démarche permet d'identifier les réussites, les faiblesses et les opportunités d'amélioration. L'ajustement des méthodes et outils utilisés devient alors essentiel pour maintenir une démarche analytique cohérente avec les objectifs fixés. Une approche flexible et évolutive reste indispensable face à un environnement dynamique et changeant [ins23, IT23].

3.4.8 Les indicateurs clés de performance (KPIs)

KPIs sont des mesures quantitatives essentielles permettant d'évaluer les avancées vers les objectifs stratégiques. Ils font le lien entre ambition et action, en jouant le rôle de repères directionnels pour l'ensemble de l'organisation.

Un KPIs efficace doit répondre à plusieurs critères essentiels :

- **Défini avec précision** : il doit être compris sans ambiguïté, avec une définition claire, une valeur cible explicite et un délai de référence bien établi.
- **Quantifiable** : il doit permettre un suivi chiffré dans le temps pour mesurer les progrès réalisés.
- **Orienté sur un objectif clair** : les cibles doivent être à la fois ambitieuses et réalistes, accessibles par des efforts concrets.
- **Visible** : les indicateurs doivent être consultables par les acteurs concernés et, idéalement, partagés avec l'ensemble du personnel afin de favoriser un retour pertinent sur les actions entreprises.
- **En cohérence avec la stratégie globale** : Il est essentiel de ne pas se limiter aux mesures faciles, mais de veiller à ce qu'elles reflètent réellement les objectifs stratégiques, comme la satisfaction client plutôt que de simples métriques opérationnelles.

Pour qu'un KPI soit smart, il doit éviter les formulations floues telles que "améliorer", "renforcer", "de qualité" ou encore "meilleur". Stacey Barr, experte en performance organisationnelle, qualifie ces termes de "mots fouines" (*weasel words*). Elle recommande de les remplacer par des objectifs précis, concrets et mesurables, comme le montre le tableau ci-dessous :

Mauvais objectifs	Bons objectifs (smart)
Améliorer la qualité de la production.	Réduire de 10 % le taux de rebuts sur la ligne d'assemblage 'A' d'ici le 31/12/2025.
Offrir un service logistique plus rapide.	Réduire le délai moyen de livraison client de 3 jours à 2 jours avant la fin de l'année
Assurer un environnement de travail sûr pour tous.	Réduire le taux de fréquence des accidents de travail avec arrêt à moins de 3 pour 1 million d'heures travaillées d'ici l'année prochaine.

TABLE 3.2 – Exemples d'objectifs smart en industrie

Il est préférable de limiter le nombre de KPIs afin de rester concentré sur l'essentiel. Une entreprise peut structurer sa stratégie autour de 4 à 5 axes (finance, marketing, technologie ...), chacun comprenant 2 à 5 objectifs stratégiques, associés à 1 à 3 KPIs. Cela représente un total recommandé de 16 à 30 indicateurs. Un excès d'indicateurs peut disperser l'attention, engendrer de la fatigue et nuire à la performance globale [And15].

3.4.9 Data culture

La data culture, ou culture des données, désigne une organisation où les décisions sont prises sur la base de données fiables, plutôt que sur l'intuition ou l'autorité hiérarchique. Construire une telle culture implique d'adopter des pratiques favorisant l'objectivité, la collaboration et l'usage systématique des faits mesurables.

Instaurer une véritable culture data-driven nécessite de limiter l’influence des décisions fondées uniquement sur l’intuition ou le statut hiérarchique. Les Highest Paid Person’s Opinion (HiPPOs) freinent l’objectivité, l’innovation et la collaboration. Pour y remédier, il faut responsabiliser les décideurs sur leurs résultats, encourager l’ouverture au doute, promouvoir la pratique des tests, et valoriser les idées appuyées par des données, indépendamment de leur origine.

Promouvoir une culture axée sur les données repose avant tout sur la confiance, l’ouverture et l’accessibilité. Les collaborateurs doivent pouvoir faire confiance non seulement à la qualité des données, mais aussi à leur usage éthique et bénéfique. Cela passe par la démocratisation de l’accès aux données, chaque membre de l’organisation doit pouvoir y accéder dans le respect du cadre légal. Cette transparence permet de décentraliser la prise de décision, en dotant les employés des bons outils et compétences, ils peuvent agir localement avec pertinence, tout en allégeant la pression sur les échelons supérieurs. Toutefois, une telle culture ne se construit pas spontanément, elle nécessite des actions volontaristes de partage, d’éducation et de gestion du changement.

La réussite de cette transformation dépend aussi de l’élargissement de la compréhension des données à tous les niveaux de l’organisation, en particulier parmi les décideurs. Sans devenir des experts techniques, les managers doivent acquérir une culture statistique de base, comprendre les principes d’inférence, d’extrapolation et de design expérimental. Cela leur permettra d’évaluer la pertinence des analyses, de reconnaître les mauvaises interprétations, et de prendre des décisions éclairées. Leur rôle est aussi crucial dans la validation des investissements en outils d’analyse, le soutien aux initiatives de formation, et la communication des résultats aux parties prenantes. À l’inverse, un déficit de littératie statistique expose l’entreprise à des décisions erronées et à une mauvaise évaluation des risques.

Une étude menée par l’Economist Intelligence Unit montre que les organisations qui surperforment sont celles dont la direction s’engage activement dans l’usage des données. Comme l’illustre la Figure (3.3), Les entreprises dont les dirigeants valorisent les données en les rendant visibles, utiles et centrales dans le processus décisionnel obtiennent généralement de meilleurs résultats que celles où l’engagement du leadership est plus faible. Cela souligne le rôle déterminant de l’exemplarité managériale dans la construction d’une culture véritablement orientée données.

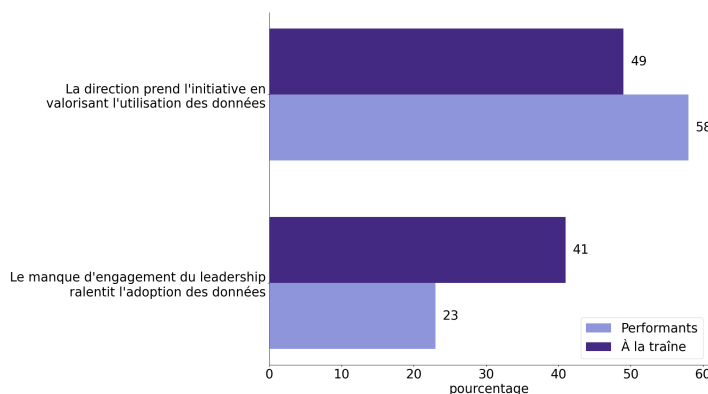


FIGURE 3.3 – Impact du leadership sur l’adoption d’une culture orientée données.

Certaines entreprises pionnières, comme Facebook, ont compris cette nécessité. Elles ont mis en place des initiatives ambitieuses telles que le "Data Camp", une formation immersive de deux semaines, ouverte non seulement aux analystes, mais aussi aux chefs de projet, designers, équipes financières, service client et opérations. L'objectif est clair, démocratiser l'usage des données dans chaque métier. Même si toutes les entreprises ne disposent pas des ressources d'un géant technologique, des solutions accessibles existent : formations gratuites en ligne (Coursera, Udacity, Khan Academy), ouvrages de référence (comme OpenIntro Statistics), ou programmes de montée en compétence adaptés à différents publics. L'essentiel est de s'engager progressivement dans cette voie pour construire une organisation durablement orientée vers les données [And15].

Enfin, construire une culture data-driven durable implique aussi d'intégrer pleinement la dimension éthique et la protection de la vie privée dans l'usage des données. La démocratisation de l'accès aux données ne doit pas se faire au détriment des droits des utilisateurs ou des clients. Une organisation responsable est celle qui respecte non seulement les exigences légales, mais qui adopte également des principes internes plus stricts en matière de confidentialité, d'intégrité des données et de gestion des risques.

La Federal Trade Commission (FTC) a défini dès 1998 cinq principes fondamentaux qui restent pleinement d'actualité pour encadrer une utilisation éthique des données [And15, Ana23] :

- **Notification/Awareness** : informer clairement les utilisateurs des pratiques de collecte avant toute utilisation de leurs données.
- **Choix/Consentement** : offrir des options réelles sur la manière dont leurs données seront utilisées.
- **Accès/Participation** : permettre aux utilisateurs d'accéder à leurs données et de corriger d'éventuelles erreurs.
- **Intégrité/Sécurité** : mettre en œuvre des mesures robustes pour sécuriser les données contre les accès non autorisés et les pertes.
- **Application/Réparation** : établir des mécanismes de contrôle et de recours pour garantir l'application de ces principes.

Respecter ces fondements n'est pas seulement une obligation morale, c'est aussi un levier stratégique, cela renforce la confiance des parties prenantes, protège la réputation de l'organisation et soutient la pérennité de la culture data-driven.

3.4.10 Les outils clés pour devenir une entreprise data-driven

Une entreprise data-driven doit s'appuyer sur un écosystème technologique cohérent, capable de soutenir l'ensemble du cycle de vie des données. De la collecte à l'analyse, en passant par le traitement et la visualisation, chaque outil joue un rôle clé dans la transformation des données brutes en actions stratégiques concrètes.

3.4.10.1 Plateformes de BI

Le marché actuel propose une grande variété de logiciels de BI permettant aux entreprises d'adopter une approche data-driven. Le tableau ci-dessous (3.3) présente trois principaux outils de BI en 2025 [Pro25] :

Logiciel	Taille d'entreprise cible	Fonctionnalités clés
Qlik Sense	Petites, moyennes et grandes entreprises.	Exploration libre des données grâce à la technologie associative, intégration d'IA et machine learning, analyse prédictive. Version gratuite disponible (Qlik Sense Desktop).
Power BI	Petites à grandes entreprises.	Intégration facile avec Microsoft, Power Query, Power Pivot, A/B Testing, personnalisation des visualisations, support de DAX, R, Python, et SQL.
Tableau	Petites à grandes entreprises.	Visualisation interactive, Data Stories avec langage naturel, partage via Tableau Server/Online, intégration avec Slack, fonctionnalités de cartographie avancée.

TABLE 3.3 – Top 3 des logiciels de BI en 2025.

3.4.10.2 Outils d'analyse prédictive

Le marché actuel propose une multitude d'outils d'analyse prédictive, chacun répondant à des besoins spécifiques en matière d'exploration de données. Le tableau ci-dessous (3.4) présente une sélection de trois solutions majeures [cli24] :

Logiciel	Fonctionnalités	Limites
SAP Analytics Cloud	Intégration SAP, planification de scénarios, visualisation de données, collaboration, contenu préétabli pour la planification.	Données parfois mal résumées, chargement lent des modèles en direct.
Amazon QuickSight	Prévisions intelligentes avec exclusion des anomalies, intégration aux dashboards, requêtes en langage naturel (Q), cloud évolutif, appli mobile.	Difficile à gérer pour grands ensembles de données, connecteurs limités.
Oracle Analytics	Visualisations riches (45 types), modèles ML personnalisés, déploiement cloud/on-prem/hybride, intégration Azure/Google Cloud, API pour intégration.	Mise en œuvre longue, temps de réponse parfois lent.

TABLE 3.4 – Les 3 meilleurs logiciels d'analyse prédictive.

3.4.10.3 CRM intégrés

Les CRM sont essentiels dans une stratégie data-driven, car ils centralisent l'ensemble des données clients en un seul point. Cette centralisation offre aux entreprises une vision complète de chaque client, favorisant une meilleure compréhension de leurs besoins et comportements, ainsi qu'une personnalisation accrue des interactions. De plus, un CRM intégré renforce la collaboration entre les équipes (ventes, marketing, service client) en leur donnant accès à des informations pertinentes et constamment mises à jour [Sal24].

Dans ce qui suit, nous présentons quelques exemples d'outils CRM intégrés (Tab 3.5) parmi les solutions les plus reconnues en 2025 [Gui25] :

Outil CRM	Description
monday.com CRM	CRM personnalisable sans code, adapté à tous les secteurs. Il permet de gérer plusieurs pipelines, d'automatiser les tâches répétitives et d'exploiter l'IA pour générer des e-mails, des transcriptions ou des tâches à partir des réunions.
Pipedrive	CRM SaaS idéal pour les PME, avec une interface glisser-déposer, des automatisations (emails, appels), plus de 500 intégrations et un chatbot actif 24h/24 pour générer des leads.
HubSpot	CRM complet avec version gratuite. Il couvre tout le cycle de vente, propose des outils d'automatisation, d'analyse, d'intelligence conversationnelle et plus de 1 500 intégrations via sa marketplace.

TABLE 3.5 – Exemples d'outils CRM intégrés.

3.4.10.4 Outils de collecte de données

Un service de collecte de données est une plateforme en ligne permettant de centraliser automatiquement des données issues de diverses sources. On distingue trois grandes catégories de ces services :

- **Le web scraping** : qui extrait automatiquement des données depuis des pages web, souvent via des proxies.
- **Les API spécialisées** : qui permettent d'accéder à des données structurées depuis diverses plateformes.
- **Les services de récupération de données** : qui agrègent des informations issues de multiples sources dans des ensembles cohérents, parfois enrichis selon les besoins.

Il convient de souligner qu'un même service peut remplir simultanément plusieurs de ces fonctions, ces catégories n'étant pas exclusives entre elles [Sal24].

Le tableau ci-dessous (3.6) présente quelques-uns des meilleurs services de collecte de données en 2025 [Dat25] :

Outils	Description
Bright Data	Propose une large gamme de solutions de collecte de données, incluant le web scraping, les API et des services de récupération de données via sa marketplace. Elle offre des ensembles de données prêts à l'emploi ou personnalisés, conformes aux normes (RGPD, CCPA), et soutenus par une équipe d'expert.
NetNut	Fournisseur réputé de proxies avec des solutions anti-bots avancées et une API efficace. Il propose également des bases de données massives, idéales pour les analyses B2B. Tous les produits sont disponibles en test gratuit.
Smartproxy	Connu pour ses proxies performants, Smartproxy fournit aussi des outils de scraping adaptés aux réseaux sociaux, à l'e-commerce et aux moteurs de recherche. Il propose une API web, des solutions pour contourner les blocages, et une interface conviviale.

TABLE 3.6 – Exemples de services de collecte de données en 2025.

3.4.10.5 Outils de gestion des données

La gestion des données joue un rôle fondamental dans la garantie de la qualité, de l'intégrité et de l'accessibilité des informations. Des solutions telles que Data Cloud permettent de structurer, nettoyer et stocker efficacement des volumes importants de données provenant de sources variées. Ces outils assurent la cohérence et la fiabilité des données, condition essentielle pour des analyses pertinentes et exploitables [Sal24].

3.4.10.6 Plateformes de Marketing Automation

Les plateformes de marketing automation jouent un rôle clé dans l'optimisation des campagnes marketing. En exploitant les données collectées, elles permettent d'automatiser l'envoi de messages personnalisés, de segmenter efficacement les audiences et de suivre les performances en temps réel. Ces outils sont essentiels pour renforcer l'impact des actions marketing tout en garantissant une expérience client cohérente et adaptée à chaque profil [Sal24].

Le tableau ci-dessous (3.6) présente quelques-uns des meilleurs services de logiciel marketing automation en 2025 [Act25] :

Plateforme	Description
ActiveTrail	Est une solution tout-en-un de marketing automation avec CRM intégré, permettant la création de scénarios multicanaux automatisés et la segmentation dynamique des contacts via une interface intuitive.
Integromat	Plateforme d'automatisation visuelle permettant de connecter de nombreuses applications et de créer des scénarios complexes, avec plus de 30 fonctionnalités. Puissante mais uniquement en anglais, elle requiert souvent des compétences en codage.
Marketo	Solution d'Adobe dédiée aux grandes entreprises, offrant un engagement omnicanal, une personnalisation avancée et des rapports détaillés. Elle se distingue par son intégration complète des canaux marketing et un support client performant.

TABLE 3.7 – Exemples de plateformes de Marketing Automation.

Au-delà des logiciels prêts à l'emploi, l'utilisation de langages de programmation comme *Python* et *R* permet de développer des analyses prédictives sur mesure. Grâce à des bibliothèques telles que *scikit-learn*, *TensorFlow* ou *Prophet*, ces langages offrent une grande flexibilité pour appliquer des méthodes avancées de ML, réaliser des analyses descriptives ou exploratoires, et générer automatiquement des rapports adaptés aux besoins spécifiques de l'entreprise.

3.5 Success Stories : L'impact du data-driven sur les entreprises populaires

3.5.1 Amazon

Amazon, fondée il y a plus de 27 ans comme librairie en ligne, est devenue un leader mondial de l'e-commerce, offrant des services tels qu'Amazon Prime et AWS. Avec plus de 150 millions d'utilisateurs, l'entreprise domine le secteur, AWS étant désormais sa principale source de profit.



Le principal défi d'Amazon a été de gérer efficacement son vaste inventaire et sa chaîne d'approvisionnement pour répondre à la demande mondiale et maintenir sa compétitivité.

Amazon a su surmonter ce défi en exploitant la puissance des données [Ana23, Jou24, Ser22], comme en témoigne la stratégie détaillée dans le tableau (3.8) suivant :

Stratégie Data-Driven	Description	Résultats
Analyses en temps réel	Amazon collecte des données en temps réel provenant de ses ventes, du comportement des consommateurs et des opérations logistiques pour prévoir la demande et ajuster les niveaux d'inventaire.	Optimisation des stocks, réduction des ruptures de stock et gestion efficace des coûts opérationnels.
Apprentissage automatique	Utilisation de modèles de ML pour prédire la demande future en se basant sur les tendances du marché, le comportement des utilisateurs et la saisonnalité.	Amélioration de la gestion des stocks et des prévisions plus précises, permettant une disponibilité continue des produits populaires.
Prix dynamiques	Mise en place de stratégies de tarification dynamique qui ajustent les prix en fonction de la demande, de la concurrence et d'autres facteurs économiques.	Maintien d'une compétitivité constante sur le marché et maximisation des revenus par une tarification adaptée.

TABLE 3.8 – Stratégies Data-Driven d'Amazon.

3.5.2 Starbucks

Starbucks, fondée en 1971, est devenue l'une des plus grandes chaînes de cafés au monde, avec des milliers de magasins dans plusieurs pays. L'entreprise est particulièrement reconnue pour sa capacité à personnaliser l'expérience client et pour son modèle d'affaires centré sur la fidélisation et l'engagement des consommateurs.



Le principal défi de Starbucks a été d'améliorer la personnalisation de l'expérience client et l'efficacité de ses campagnes marketing afin de répondre aux attentes de ses clients tout en maximisant son retour sur investissement.

Starbucks a relevé ce défi en exploitant les données des clients à travers son programme de fidélité et en mettant en œuvre des campagnes marketing ciblées [Ana23, Jou24, Ser22, Ber22], comme en témoigne la stratégie détaillée dans le tableau (3.9) suivant :

Stratégie Data-Driven	Description	Résultats
Programme de fidélité personnalisé	Starbucks collecte des données sur les achats et préférences des clients via son programme de fidélité, afin de proposer des récompenses et promotions personnalisées.	Augmentation de l'engagement des clients, avec un nombre accru de transactions récurrentes et une contribution importante aux revenus globaux.
Marketing ciblé	Utilisation de l'analyse prédictive pour personnaliser les campagnes marketing et les promotions en fonction des comportements et préférences des clients.	Amélioration des taux de conversion et de la satisfaction client, optimisant ainsi les efforts marketing et renforçant la fidélité.
Optimisation des emplacements	Grâce aux données collectées via l'application mobile et les points de vente, Starbucks évalue les habitudes de consommation et optimise ses choix d'emplacement de magasins.	Meilleure performance des nouveaux magasins, avec des ventes supérieures à celles obtenues par des méthodes traditionnelles de sélection des sites.

TABLE 3.9 – Stratégies Data-Driven de Starbucks.

3.5.3 IBM

IBM, fondée en 1911, est une entreprise technologique mondiale qui se distingue par son expertise en IA et en analytique de données. L'entreprise a su intégrer ces technologies pour révolutionner ses opérations internes et offrir des solutions innovantes à ses clients dans des secteurs variés tels que la santé, la finance et le service client.



Le principal défi d'IBM a été d'intégrer l'IA et l'analytique dans ses processus d'affaires et de fournir des solutions IA-driven efficaces pour ses clients tout en optimisant ses propres opérations internes.

IBM a relevé ce défi en intégrant des solutions basées sur l'IA et l'analytique dans ses processus et solutions clients [Ana23], comme en témoigne la stratégie détaillée dans le tableau (3.10) suivant :

Stratégie Data-Driven	Description	Résultats
Solutions alimentées par l'IA	IBM utilise sa plateforme Watson, offrant des capacités avancées d'analytique et d'informatique cognitive pour permettre aux entreprises de prendre des décisions éclairées basées sur les données.	Innovation et efficacité accrues dans divers secteurs, notamment la santé, la finance et le service client, grâce à des décisions basées sur des analyses de données.
Optimisation des opérations internes	IBM utilise l'analytique des données pour optimiser ses propres opérations, incluant les prévisions de ventes, l'allocation des ressources et l'analyse de marché.	Amélioration de l'efficacité opérationnelle, meilleure gestion des ressources et prise de décision stratégique plus rapide.
Solutions pour les clients	IBM propose des solutions IA-driven et d'analytique à ses clients pour améliorer leurs opérations, leur prise de décision et stimuler l'innovation.	Accélération de l'innovation chez les clients, avec des résultats tangibles dans la prise de décision et l'optimisation des processus métier.

TABLE 3.10 – Stratégies Data-Driven d'IBM.

3.5.4 Netflix

Netflix, fondée en 1997, est un leader mondial du streaming de contenu, offrant une large variété de films, séries et documentaires à ses abonnés. L'entreprise utilise des algorithmes de recommandation sophistiqués pour personnaliser l'expérience utilisateur et maximiser l'engagement tout en réduisant le taux de désabonnement.

NETFLIX

Le principal défi de Netflix a été d'améliorer l'engagement des utilisateurs et de réduire le taux de désabonnement en fournissant des recommandations de contenu personnalisées et pertinentes.

Netflix a relevé ce défi en mettant en place un algorithme de recommandation basé sur les données des utilisateurs et les modèles de visionnage [Ana23], comme en témoigne la stratégie détaillée dans le tableau (3.11) suivant :

Stratégie Data-Driven	Description	Résultats
Collecte de données	Netflix recueille une grande quantité de données sur les interactions des utilisateurs, y compris l'historique des visionnages, les recherches et les évaluations.	Amélioration de la personnalisation des recommandations en se basant sur des données précises du comportement des utilisateurs.
Développement d'algorithmes	Netflix utilise des algorithmes de ML pour analyser les données et identifier des modèles et préférences. L'approche de filtrage collaboratif prédit ce que les utilisateurs pourraient apprécier en fonction du comportement des utilisateurs similaires.	Recommandations personnalisées efficaces, favorisant un engagement plus élevé des utilisateurs.
Tests A/B	Netflix effectue en continu des tests A/B sur différentes stratégies de recommandation pour optimiser les suggestions de contenu et l'interface utilisateur.	Amélioration continue des recommandations, augmentant l'engagement et la satisfaction des utilisateurs.

TABLE 3.11 – Stratégies Data-Driven de Netflix.

Conclusion

Ce chapitre a permis d'approfondir la méthodologie data-driven en présentant ses piliers fondamentaux, ses prérequis techniques et humains, ainsi que les outils indispensables à sa mise en œuvre. Il met en évidence que devenir une organisation pilotée par les données ne se limite pas à une simple digitalisation, mais implique une refonte globale de la culture, des processus décisionnels et des systèmes d'information.

L'analyse des success stories démontre que les entreprises ayant su structurer et exploiter efficacement leurs données bénéficient d'un avantage compétitif notable. Ces enseignements théoriques et pratiques serviront de base à l'étude de cas du chapitre suivant, où la transformation data-driven sera contextualisée au sein de l'entreprise Ovitala, à travers des actions concrètes d'optimisation des processus métiers.

4

Étude de cas : Optimisation des processus métiers chez Ovitale

Introduction

Ce quatrième chapitre met en application les concepts théoriques et méthodologiques abordés précédemment à travers une étude de cas concrète menée chez Ovitale, une entreprise spécialisée dans la production et la distribution d'eau minérale. Face à des enjeux liés à la gestion de la demande, des stocks et de l'approvisionnement, l'objectif est de montrer comment une approche data-driven peut améliorer la performance opérationnelle.

Pour ce faire, différentes analyses ont été menées, descriptive pour explorer les données historiques, diagnostique pour comprendre les causes des variations, prédictive pour anticiper la demande, et prescriptive pour formuler des recommandations concrètes. Ces étapes ont été accompagnées de tableaux de bord interactifs facilitant la prise de décision.

Le chapitre s'organise autour de la présentation de l'entreprise, de l'identification des besoins, de la collecte et du traitement des données, jusqu'à l'analyse et la visualisation des résultats via un pipeline analytique Python–Power BI. Il prolonge ainsi la réflexion des chapitres précédents en illustrant les bénéfices concrets d'une stratégie orientée données dans un contexte industriel réel.

4.1 Présentation de l'entreprise Ovitale

Ovitale est une marque algérienne d'eau de source embouteillée, développée par la SARL BOISSONS ITHRI, une entreprise de taille moyenne. Forte de 28 années d'expérience, cette société s'est spécialisée dans la production et la distribution d'eau embouteillée, en s'appuyant sur des valeurs de qualité, d'innovation et de proximité avec le marché national. Son siège social est implanté à Akbou, dans la wilaya de Béjaïa, sur la route du Marché de Gros.



FIGURE 4.1 – Logo d’Ovitale.

L’entreprise a connu une évolution progressive, marquée par des étapes clés, comme l’illustre le tableau ci-dessous (4.1) :

Année	Événement
1991	Création de l’établissement HAMIDOUCHE, dédié à la production de boissons gazeuses.
1998	Acquisition d’un terrain industriel à Akbou pour faire face à la croissance.
2000	Transformation en SARL ITHRI SODA, poursuivant l’activité de production de boissons.
2003	Suite à la découverte d’une nappe phréatique, la société devient SARL BOISSONS ITHRI, amorçant une diversification vers l’eau embouteillée.
2012	Lancement officiel de la marque Ovitale, avec un premier format de 5 litres d’eau de source.
2015	Extension de la gamme de produits avec les formats 1.5 Litre, 0.5 Litre, et 0.5 Litre bouchon sport.
2018	Arrêt définitif de la fabrication de sodas, afin de se concentrer exclusivement sur l’eau embouteillée.
2019	Mise à jour des lignes de production avec l’acquisition de machines à haute capacité, conformes aux normes internationales. Cette évolution marque l’entrée progressive de l’entreprise dans l’industrie 4.0, renforçant son efficacité et sa compétitivité industrielle.
2025	Certification ISO 22000, assurant la mise en place d’un système de gestion efficace de la sécurité des aliments, conforme aux normes internationales et garantissant des produits sûrs pour la consommation.

TABLE 4.1 – Historique d’Ovitale

Le site industriel d'Oviale s'étend sur une superficie de 8 hectares et comprend plusieurs unités de production spécialisées selon les formats (Fig 4.2) :

- Unité 1.5 Litre.
- Unité 0.5 Litre.
- Unité 6 Litre.



FIGURE 4.2 – Les trois formats de bouteilles produits par Ovitale.

Grâce à son réseau logistique structuré, OVITALE distribue ses produits dans 35 wilayas algériennes, consolidant une présence régionale forte et une reconnaissance croissante dans le secteur de l'eau embouteillée [Ovi].

4.2 Analyse de besoins

4.2.1 Diagnostic initial

Dans le cadre de notre stage chez Ovitale, nous avons entamé notre mission par une immersion sur le terrain. Cette phase exploratoire nous a permis de visiter les différents services de l'entreprise et d'observer de près les principaux processus métiers en place.

Au fil de cette exploration, nous avons constaté que, bien que l'entreprise génère quotidiennement un volume important de données, celles-ci ne sont pas réellement exploitées à des fins décisionnelles. Les décisions, qu'elles soient stratégiques ou opérationnelles, semblent davantage guidées par l'expérience ou l'intuition, plutôt que par une analyse structurée et objective des données disponibles. Cette absence de pilotage par les données limite considérablement la capacité d'Ovitale à optimiser ses processus et à anticiper les évolutions du marché.

4.2.2 Définition des objectifs stratégiques

À la suite du diagnostic initial, nous avons proposé une approche data-driven et défini ses objectifs en collaboration avec la direction générale de l'entreprise, dans le but de tirer pleinement parti des données disponibles pour optimiser les processus métiers de l'organisation à travers trois axes majeurs :

- **Stratégie commerciale orientée données** : mettre en place un système de prévision des ventes afin d’anticiper la demande, ajuster les volumes de production et renforcer la prise de décision commerciale.
- **Approvisionnement intelligent** : concevoir un système de recommandations facilitant les décisions d’achat, en prenant en compte les tendances de consommation, les capacités de la production.
- **Optimisation de la gestion des stocks** : mettre en place une optimisation proactive des stocks de PF et de MP dans le but de limiter les surstocks, éviter les ruptures, et assurer une meilleure synchronisation entre la production et l’approvisionnement.

4.3 Collecte des données

À la suite de la phase d’analyse des besoins et de la définition des objectifs stratégiques, nous avons entamé l’étape de collecte des données, une étape essentielle à la mise en œuvre d’une approche data-driven. Elle a consisté à réunir un ensemble de données internes et externes.

4.3.1 Données internes

La collecte des données internes s’est appuyée sur les données générées par différents services clés de l’entreprise, sous forme de fichiers Excel et CSV, couvrant principalement la période allant de 2020 à 2024, comme le montre le tableau ci-dessous (4.2) :

Service concerné	Données collectées
Service commercial	Données journalières de ventes, incluant les quantités vendues par format de bouteille.
Service de production	Données mensuelles de production ventilées par format de bouteille, capacité de production (cadence des machines en unités/heure), et coût de lancement d’un lot de production.
Service d’approvisionnement	Données relatives aux achats et à la consommation des MP.
Service de gestion des stocks	Mouvements mensuels de stock pour les produits finis et les MP, ainsi que le coût de stockage par unité.

TABLE 4.2 – Sources internes de données utilisées dans le projet.

4.3.2 Données externes (contextuelles)

En complément des données internes, un ensemble de données externes, collectées à partir de 2020, a été rassemblé. Ces données, susceptibles d’influencer les ventes, sont présentées dans le tableau ci-dessous (4.3) :

Outil utilisé	Données collectées
Web scraping	Événements saisonniers, culturels et sportifs (marathons internationaux, festivals culturels, compétitions régionales) dans différentes Wilayas d'Algérie.
API Open-Meteo	Données météorologiques (températures, taux d'humidité) des 58 wilayas d'Algérie.
Web scraping	Périodes particulières du calendrier (week-ends, jours fériés, mois de Ramadan).

TABLE 4.3 – Sources et outils de collecte des données externes.

4.4 Traitement des données pour l'analyse

Avant de procéder aux différentes analyses, une étape de traitement des données est essentielle. Elle consiste à nettoyer, transformer et structurer les données collectées, afin de garantir leur qualité, leur cohérence et leur exploitabilité.

Concrètement, ce processus a impliqué plusieurs opérations fondamentales, notamment :

- La suppression des doublons.
- La vérification et l'uniformisation des types de données.
- La mise à l'échelle des unités (afin d'assurer l'harmonisation entre les variables).
- la gestion des valeurs manquantes et aberrantes.

Ce traitement a permis de répondre aux exigences des quatre types d'analyses menées dans ce projet :

- **L'analyse descriptive** : elle répond à la question « Que s'est-il passé ? ». Elle consiste à explorer et comprendre les données historiques de chaque service. Elle met en évidence des tendances saisonnières, des variations récurrentes ainsi que des anomalies potentielles, offrant ainsi une première lecture structurée des performances de l'entreprise. Pour ce faire, des tableaux de bord interactifs sont mis en place afin de faciliter le suivi opérationnel au quotidien.
- **L'analyse diagnostique** : elle vise à répondre à la question « Pourquoi cela s'est-il produit ? ». Elle analyse les corrélations entre les fluctuations des ventes et divers facteurs explicatifs, qu'ils soient internes ou contextuels. Cette approche permet de mieux cerner les causes sous-jacentes aux variations observées et de fonder les décisions sur des éléments objectifs et vérifiables.
- **L'analyse prédictive** : elle s'attache à répondre à la question « Que va-t-il se passer ? ». Elle repose sur l'utilisation de modèles de ML pour anticiper l'évolution future de la demande. Les résultats sont présentés sous forme de visualisations interactives, facilitant ainsi leur interprétation et leur exploitation dans les processus de décision.
- **L'analyse prescriptive** : elle répond à la question « Que devons-nous faire ? ». En se basant sur les résultats des prévisions, elle propose des recommandations concrètes pour

ajuster la production, optimisé la gestion des stocks ainsi que l’approvisionnement. Elle s’appuie également sur des visualisations interactives et sur des algorithmes d’optimisation pour appuyer les choix stratégiques.

Dans les sections suivantes, nous présenterons en détail la démarche adoptée pour chaque type d’analyse, en mettant en évidence les résultats obtenus, leur interprétation ainsi que les recommandations qui en découlent.

4.5 Analyse descriptive

À l’issue de la phase de préparation et de nettoyage, les données ont été modélisées dans Power BI selon une architecture en constellation (Fig4.3). Cette structure garantit une intégrité référentielle robuste, assure la cohérence des données et offre une grande flexibilité pour la réalisation d’analyses multidimensionnelles et interactives. Elle repose sur :

- **Les tables de faits** : représentant les processus métiers, elle contiennent des mesures quantitatives ainsi que les clés étrangères permettant de les relier aux dimensions concernées.
- **Les tables de dimension** : regroupant les axes d’analyse, elles contiennent les clés primaires ainsi que des attributs descriptifs facilitant l’interprétation des faits.

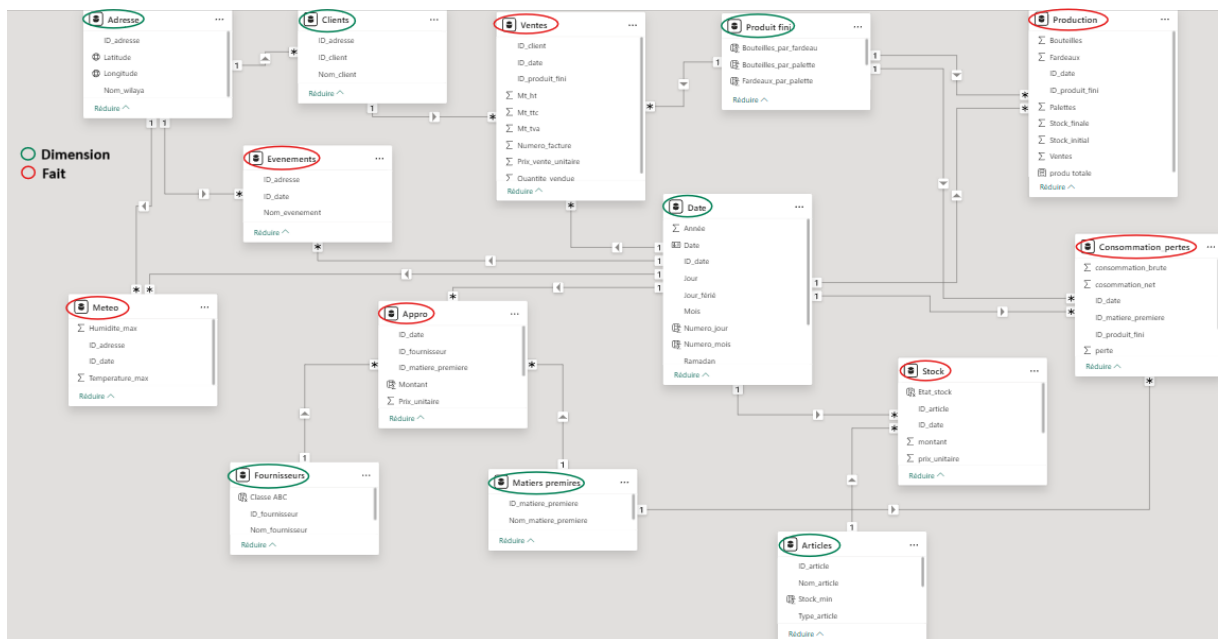


FIGURE 4.3 – Schéma en constellation 1.

À partir de cette architecture en schéma en constellation, nous avons développé des tableaux de bord interactifs sous Power BI pour visualiser les données.

4.5.1 Analyse descriptive des ventes

Ce premier tableau de bord (Fig 4.4) a été conçu pour assurer un suivi détaillé des ventes sur la période 2020–2024. Il propose une interface interactive qui permet d’explorer les données à travers plusieurs dimensions temporelles et commerciales.

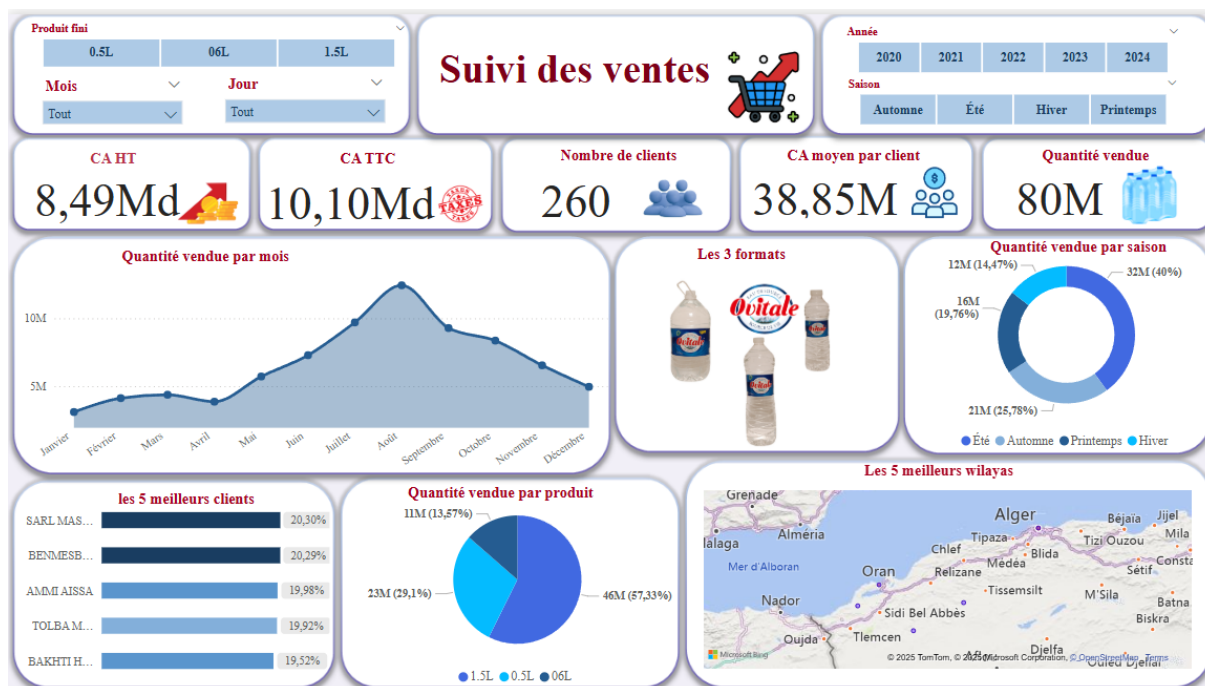


FIGURE 4.4 – Suivi des ventes.

Analyse des KPIs

Afin d’évaluer la performance commerciale de l’entreprise sur la période 2020–2024, plusieurs KPIs ont été extraits et analysés. Ces KPIs offrent une vision synthétique de l’évolution des ventes, du chiffre d’affaires (CA) et de la base clients. Le tableau suivant (4.4) résume les principales métriques observées durant ces cinq années :

Année	CA HT	CA TTC	Nb clients	CA moyen par client	Quantité vendue
2020	926,55 M	1,10 Md	95	11,61 M	10 M
2021	1,33 Md	1,58 Md	85	18,57 M	13 M
2022	1,73 Md	2,06 Md	97	21,19 M	16 M
2023	2,10 Md	2,50 Md	102	24,49 M	18 M
2024	2,41 Md	2,87 Md	109	26,32 M	22 M
2020–2024	8,49 Md	10,10 Md	260	38,85 M	80 M

TABLE 4.4 – Évolution des KPIs des ventes (2020–2024).

Durant les cinq années observées, l'entreprise a connu une croissance continue de ses performances commerciales. Le chiffre d'affaires hors taxe (CA HT) et Chiffre d'affaires toutes taxes comprises (CA TTC) a progressé d'année en année, reflétant une augmentation régulière des ventes.

Le nombre de clients a également augmenté de manière constante, ce qui montre un élargissement de la base clientèle. En parallèle, La CA moyen par client a connu une nette amélioration, traduisant un panier d'achat plus élevé ou une fidélisation plus forte.

La quantité vendue a suivi une courbe ascendante sur l'ensemble de la période, ce qui confirme une dynamique commerciale positive et une demande croissante sur les produits proposés.

Analyse de la quantité vendue par mois et par saison

Pour analyser l'évolution des ventes mensuelles sur les cinq dernières années, les quantités vendues ont été agrégées par mois et par saison, puis représentées à l'aide de courbes annuelles et d'un graphique en anneau illustrant la répartition saisonnière, comme le montre la figure ci-dessous (4.5) :

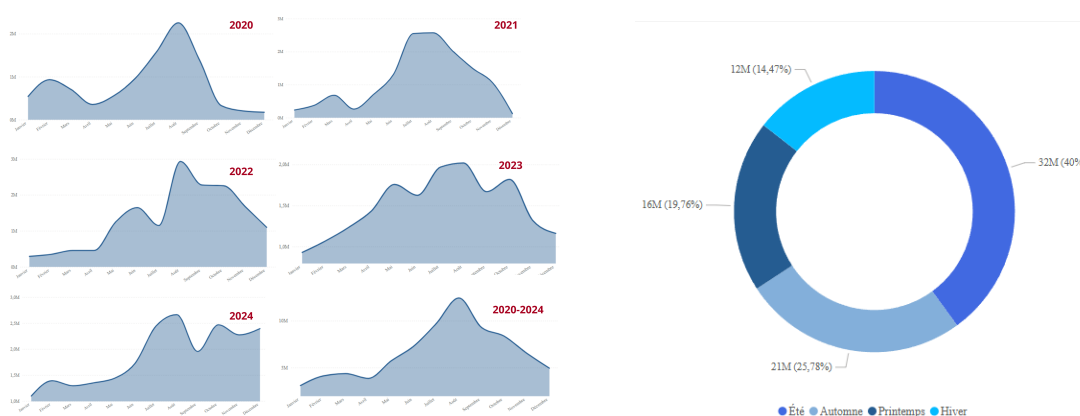


FIGURE 4.5 – Évolution mensuelle et saisonnière des ventes (2020–2024).

L'analyse des quantités vendues mensuellement entre 2020 et 2024 révèle de fortes tendances saisonnières :

- Pic des ventes en été, notamment en juillet et août, avec un maximum atteint en août (plus de 12,4 millions d'unités sur 5 ans).
- Creux des ventes en hiver, en particulier en janvier et décembre, où la consommation est nettement plus faible.
- Les mois intermédiaires présentent des variations modérées, avec une tendance à l'anticipation des pics estivaux.
- Progression globale des ventes d'année en année, traduisant une croissance soutenue de la demande.

En somme, les données confirment une saisonnalité estivale marquée et une hausse constante des volumes vendus.

Analyse de la quantité vendue par type de PF

Afin de mieux comprendre la répartition des ventes selon les formats de PF, les quantités vendues ont été agrégées par année et par format, puis représentées à l'aide de graphiques en secteurs annuels ainsi que d'un graphique global couvrant l'ensemble des cinq années, comme le montre la figure ci-dessous (4.6) :

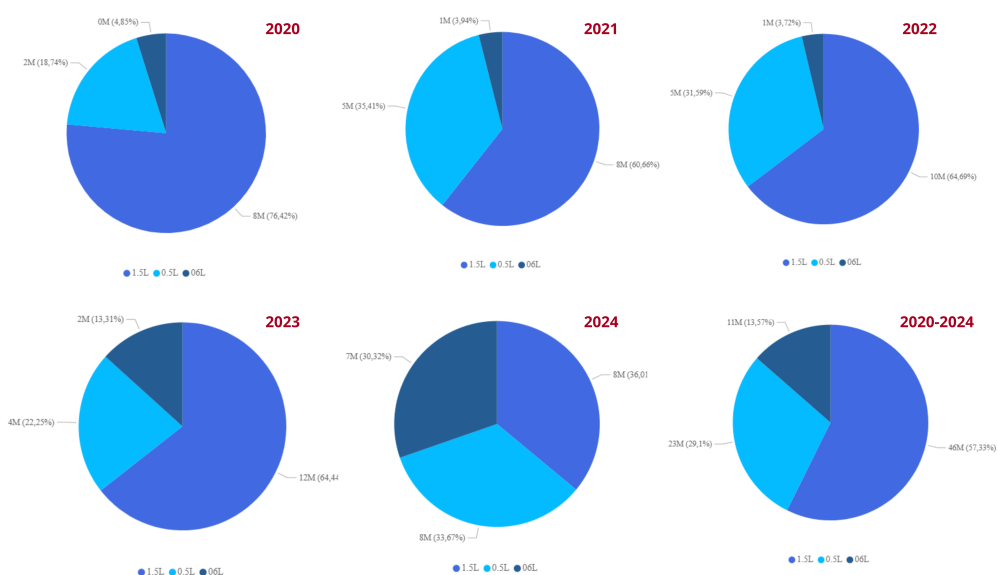


FIGURE 4.6 – Répartition des ventes par format de PF (2020–2024).

L'analyse met en évidence une prédominance marquée du format 1,5 L, qui constitue 57,33% des quantités totales vendues au cours des cinq dernières années. Il s'agit du format le plus distribué, avec un volume de 46 millions d'unités.

Le format 0,5 L arrive en seconde position avec 29,1 % des ventes, soit un total de 29 millions d'unités écoulées.

Enfin, le format 6 L représente 13,57 % des ventes, correspondant à 11 millions d'unités vendues. Ce format reste donc le moins vendu des trois.

Ces tendances révèlent une diversification croissante des préférences clients, ce qui peut inciter l'entreprise à adapter ses stratégies de production et de distribution en conséquence.

Analyse des clients stratégiques

Afin de mieux identifier les clients les plus importants, les quantités vendues ont été agrégées par année et par client, puis représentées à l'aide de graphiques en barres annuels ainsi que d'un graphique global couvrant l'ensemble des cinq années, comme le montre la figure ci-dessous (4.7) :

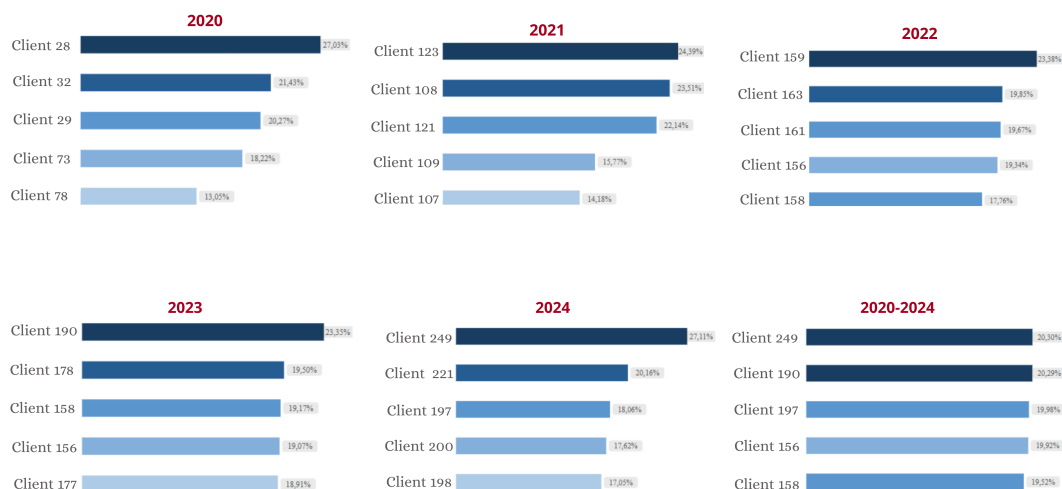


FIGURE 4.7 – Répartition des quantités vendues par client (Top 5) (2020–2024).

L’analyse annuelle révèle une variabilité des meilleurs clients d’une année à l’autre. Chaque année présente un classement légèrement différent, indiquant des changements dans les volumes commandés ou des fluctuations dans l’activité des clients.

Au cours des cinq dernières années, cinq clients se sont distingués par leur régularité d’achat et le volume important de leurs commandes. Ces clients sont désormais considérés comme des partenaires clés et figurent parmi les meilleurs clients de l’entreprise.

Bien que seuls les cinq principaux clients soient présentés dans cette analyse, les résultats révèlent qu’un nombre significatif d’autres clients réalisent également des volumes d’achats importants. Cette observation met en lumière l’existence d’une base clientèle plus large. L’entreprise dispose ainsi d’un portefeuille client solide, favorisant une répartition plus équilibrée des ventes et offrant une flexibilité commerciale accrue.

Analyse des zones géographiques à forte demande

Afin de mieux identifier les wilayas les plus importantes en termes de volume d’achats, les quantités vendues ont été agrégées par année et par wilaya. La répartition des cinq premières wilayas pour chaque année a été représentée à l’aide de cartes individuelles, accompagnées d’une carte synthétique couvrant l’ensemble des cinq années, comme le montre la figure ci-dessous (4.8) :

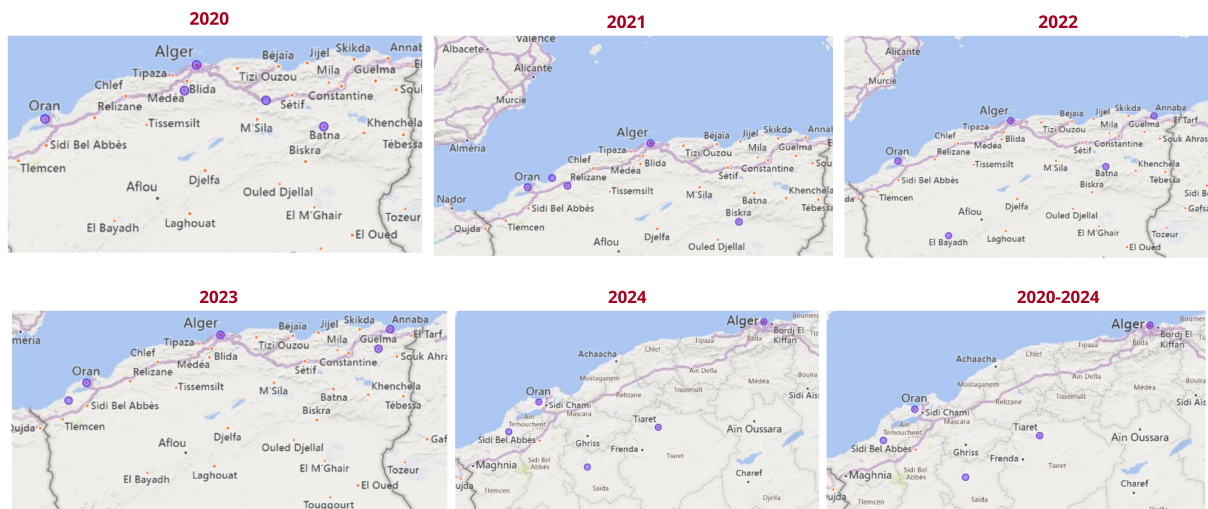


FIGURE 4.8 – Répartition géographique des ventes par wilaya (Top 5) (2020-2024).

L’analyse annuelle révèle une variabilité des meilleures wilayas d’une année à l’autre. Chaque année présente un classement légèrement différent, indiquant une dynamique commerciale évolutive selon les régions et des fluctuations dans la demande locale.

Cependant, sur l’ensemble de la période 2020–2024, cinq wilayas se distinguent par leur régularité et le volume significatif des quantités achetées (Oran, Alger, Tiaret, Aïn Témouchent et Saïda). Leur présence récurrente dans le classement annuel met en évidence leur rôle stratégique dans la performance commerciale de l’entreprise à l’échelle nationale.

4.5.2 Analyse descriptive de la production

Ce deuxième tableau de bord (Fig 4.9) a été élaboré pour assurer un suivi opérationnel de la production sur la période 2020–2024. Il propose une interface interactive qui permet d’explorer les données à travers plusieurs dimensions (temporelle, MP, formats de produits et unités de mesure).

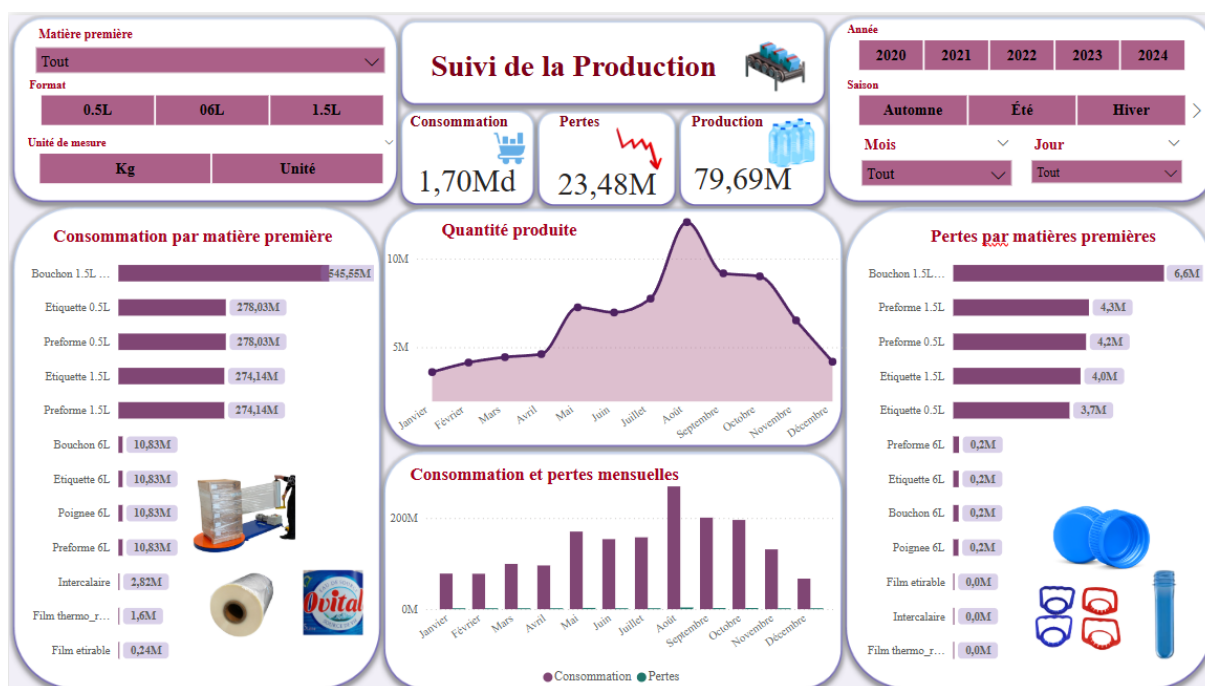


FIGURE 4.9 – Suivi de la production.

Analyse des KPIs

Afin d'évaluer l'efficacité opérationnelle de l'entreprise sur la période 2020–2024, plusieurs KPIs relatifs à la production ont été extraits et analysés. Ces indicateurs permettent de suivre l'évolution des volumes produits, la consommation des MP ainsi que les pertes enregistrées lors du processus de fabrication. Le tableau suivant (4.5) présente une synthèse des principales métriques observées sur les cinq dernières années :

Année	Consommation MP	Pertes MP	Production
2020	208,23 M	3,99 M	10,08 M
2021	318,19 M	10,31 K	13,34 M
2022	373,81 M	4,94 M	16,20 M
2023	359,71 M	6,47 M	17,95 M
2024	437,93 M	8,07 M	22,12 M
2020–2024	1,70 Md	23,48 M	79,69 M

TABLE 4.5 – Évolution des KPIs de production (2020–2024).

Sur la période 2020–2024, l'entreprise a connu une amélioration progressive de sa performance industrielle. La consommation de MP a suivi une tendance haussière, traduisant une augmentation de l'activité de production. Les pertes ont été globalement maîtrisées, témoignant d'un bon

contrôle des processus. La production a, quant à elle, affiché une progression constante, confirmant l'efficacité croissante des opérations industrielles et la capacité de l'entreprise à répondre à une demande soutenue.

Analyse de la quantité produite par mois

Pour analyser l'évolution de la production mensuelle sur les cinq dernières années, les quantités produites ont été agrégées par mois, puis visualisées sous forme de courbes représentant chaque année. comme le montre la figure ci-dessous (4.10) :

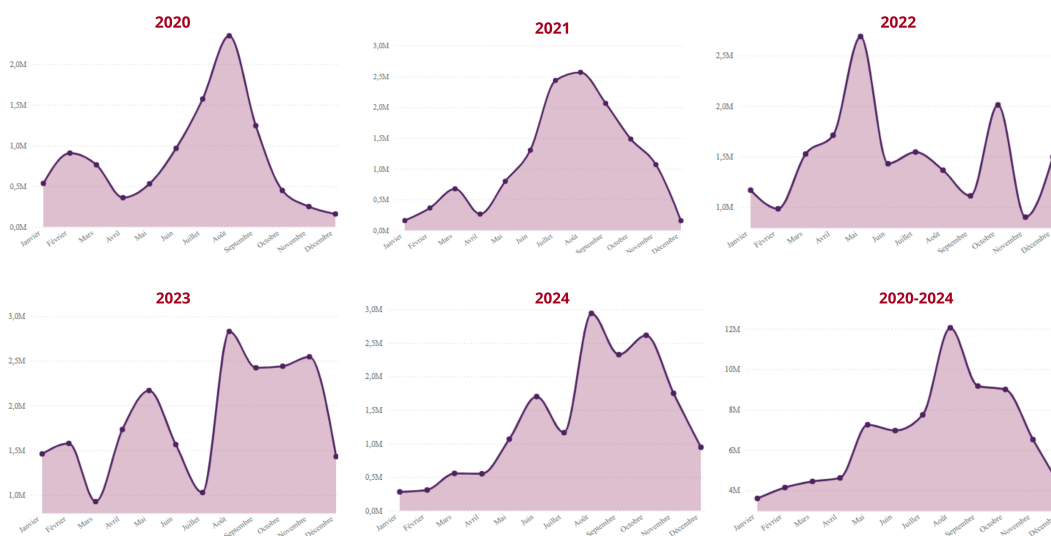


FIGURE 4.10 – Évolution mensuelle de la production (2020–2024).

L'analyse met en évidence une croissance continue de la production mensuelle au fil des années. Cette évolution s'explique en partie par les perturbations initiales dues à la pandémie de Covid-19. En 2020, les périodes de confinement ont fortement ralenti l'activité, entraînant des niveaux de production particulièrement bas. À partir de 2021, on observe une reprise graduelle, puis une accélération soutenue, reflet d'une adaptation opérationnelle et d'un rattrapage post-crise.

L'analyse met également en évidence une forte saisonnalité, avec des pics récurrents en été (notamment en août), période où la demande semble atteindre son apogée, tandis que les mois d'hiver affichent systématiquement des volumes plus faibles. On note par ailleurs des variations interannuelles dans les cycles de production, certaines années connaissant une intensification dès le printemps, d'autres concentrant l'activité en fin d'année. Sur l'ensemble de la période, l'été ressort comme une phase stratégique, appelant à une gestion rigoureuse des ressources et une planification adaptée aux pics de demande.

Analyse de la consommation et des pertes de MP

Afin d'évaluer l'efficacité de la gestion des MP, il est essentiel d'analyser leur consommation ainsi que les pertes enregistrées tout au long de l'année. Cette section présente des histogrammes

groupés illustrant, pour chaque mois, l'évolution de la consommation et des pertes sur les cinq dernières années, comme le montre la figure ci-dessous (4.11) :



FIGURE 4.11 – Évolution mensuelle de la consommation et des pertes de MP (2020–2024).

La consommation des MP montre une hausse régulière au fil des années, avec des pics saisonniers en été, notamment en août. Les mois d'hiver affichent des niveaux plus faibles, ce qui reflète une saisonnalité marquée. Malgré l'augmentation des volumes, les pertes restent globalement maîtrisées et proportionnelles à la consommation. Enfin, quelques variations interannuelles apparaissent, traduisant des ajustements ponctuels dans la planification ou les conditions de production.

Afin d'approfondir l'analyse de la consommation et des pertes des MP, nous présentons deux graphiques empilés couvrant l'ensemble de la période observée. Ils illustrent respectivement la consommation et les pertes cumulées de chaque MP, comme le montre la figure ci-dessous (4.12) :

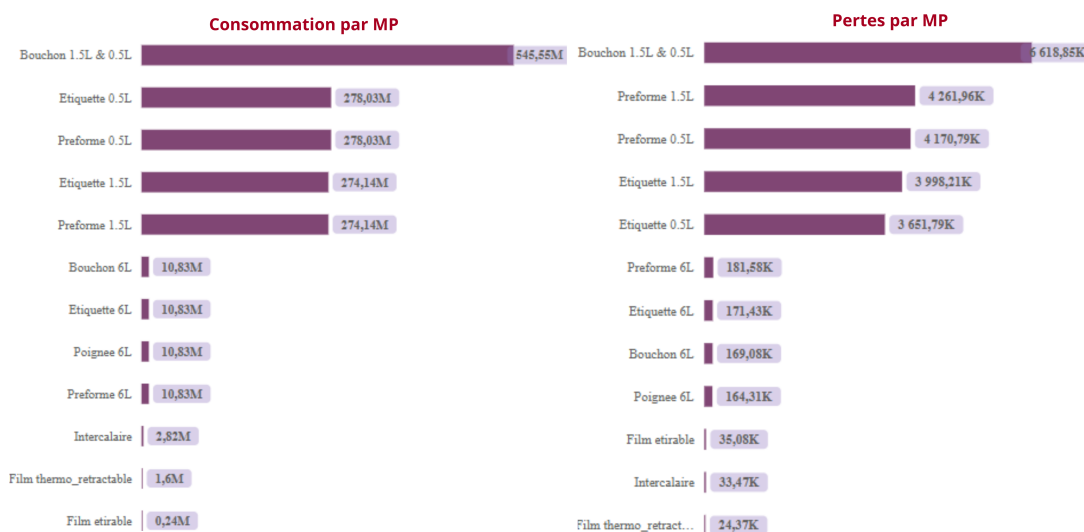


FIGURE 4.12 – Consommation et pertes cumulées des MP (2020–2024).

L’analyse de la consommation et des pertes des MP sur cinq ans révèle une forte concentration sur les formats 0,5 L et 1,5 L, notamment pour les bouchons, étiquettes et préformes, qui représentent la majorité des volumes utilisés. Les formats 6 L et les éléments auxiliaires (films, poignées, intercalaires) restent marginaux en termes de consommation.

Les pertes suivent la même tendance, elles sont plus élevées pour les MP les plus utilisées, mais proportionnelles à la consommation, ce qui reflète une gestion maîtrisée des déchets malgré l’augmentation des volumes.

4.5.3 Suivi de stock

Afin d’assurer un suivi efficace des stocks, deux tableaux de bord ont été développés pour répondre aux besoins opérationnels des gestionnaires. Le premier (4.13) permet de suivre les quantités physiques en stock, tandis que le second (4.14) se concentre sur la valorisation monétaire des stocks. Ces tableaux de bord disposent d’une interface interactive, offrant la possibilité d’explorer les données selon plusieurs dimensions (type d’article, unité de mesure, article, période), ce qui facilite leur utilisation au quotidien.

Cependant, leur structure actuelle demeure principalement orientée vers un usage opérationnel que stratégique. Leur interprétation à un niveau global reste limitée, mais cela s’avère bénéfique pour le gestionnaire de stock, qui peut ainsi contrôler et gérer les processus en temps réel de manière efficace.

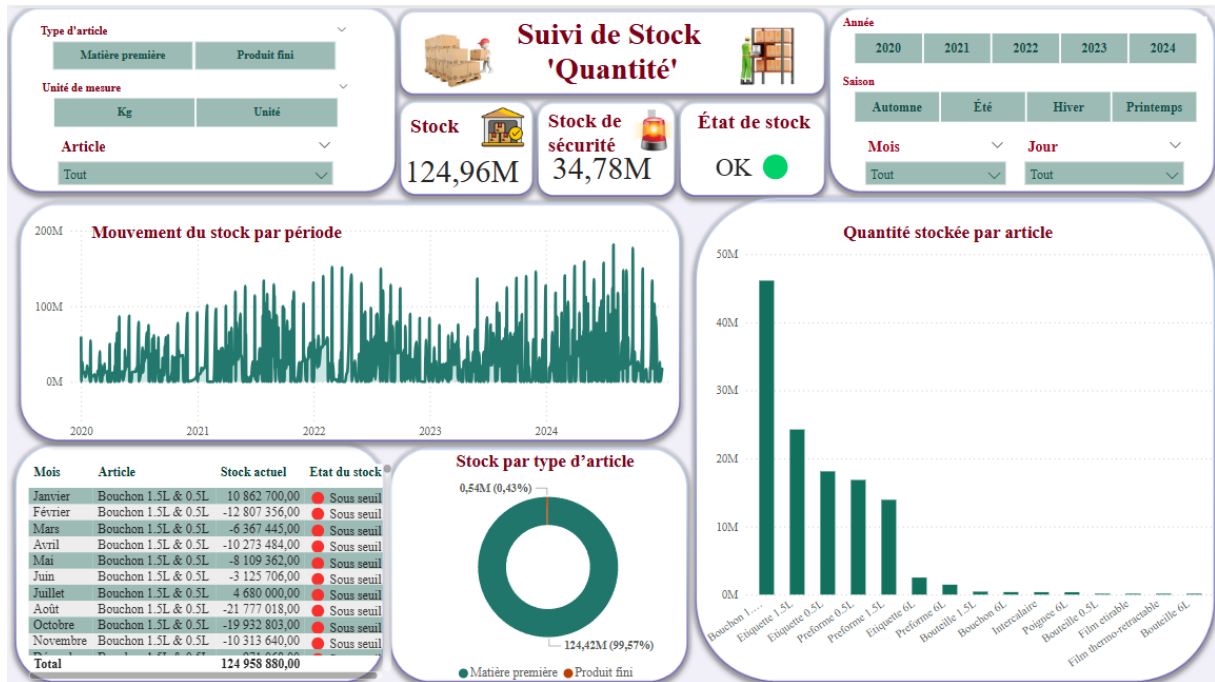


FIGURE 4.13 – Suivi de stock – Quantité.

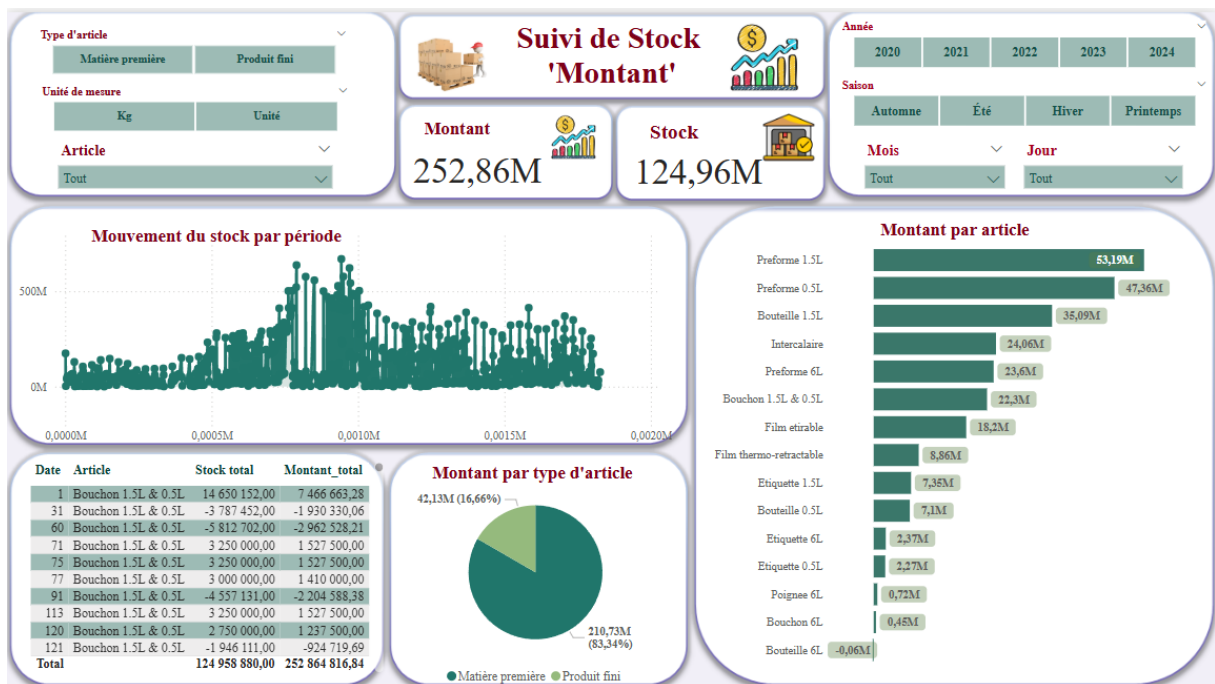


FIGURE 4.14 – Suivi de stock – Montant

4.5.4 Analyse descriptive du l’approvisionnement

Afin d’assurer un suivi détaillé de l’approvisionnement durant la période 2020–2024, un tableau de bord interactif (4.15) a été développé. Il permet d’explorer les données selon plusieurs dimensions temporelles et logistiques, comme le montre la figure ci-dessous :

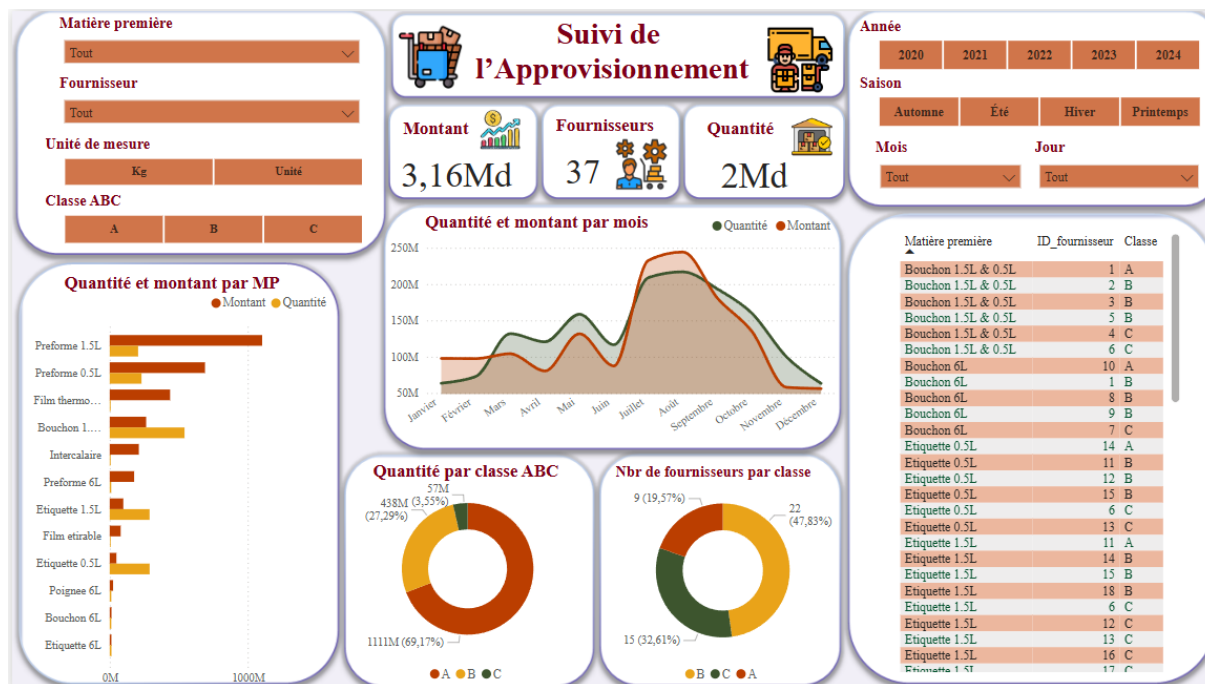


FIGURE 4.15 – Suivi de l’approvisionnement.

Analyse des KPIs

Afin d’évaluer la performance du processus d’approvisionnement entre 2020 et 2024, trois indicateurs clés ont été analysés :

Année	Quantité achetée	Montant	Nbr fournisseurs
2020	240 M	354,14 M	12
2021	308 M	693,91 M	15
2022	285 M	769,48 M	14
2023	383 M	692 M	22
2024	391 M	648,54 M	25
2020–2024	2 Md	3,16 Md	37

TABLE 4.6 – Évolution des KPIs d’approvisionnement (2020–2024).

Sur cette période, l’entreprise a acheté un total de 2 milliards d’unités de MP pour un montant global de 3,16 milliards. La quantité achetée a globalement augmenté, traduisant un renforcement de l’activité industrielle. Le nombre de fournisseurs est passé de 12 à 25, illustrant une volonté de diversification et de sécurisation des sources d’approvisionnement.

L’évolution des montants, notamment le pic observé en 2022, peut s’expliquer par une hausse des prix des MP ou par une modification dans la structure des achats. Globalement, ces indicateurs montrent un approvisionnement en croissance et de plus en plus structuré.

Analyse de la quantité achetée et du montant par mois

Afin d’assurer un suivi détaillé de la dynamique mensuelle des achats sur la période 2020–2024, les quantités achetées ainsi que les montants associés ont été analysés mois par mois, puis visualisées sous forme de courbes représentant chaque année, comme le montre la figure ci-dessous (4.16) :

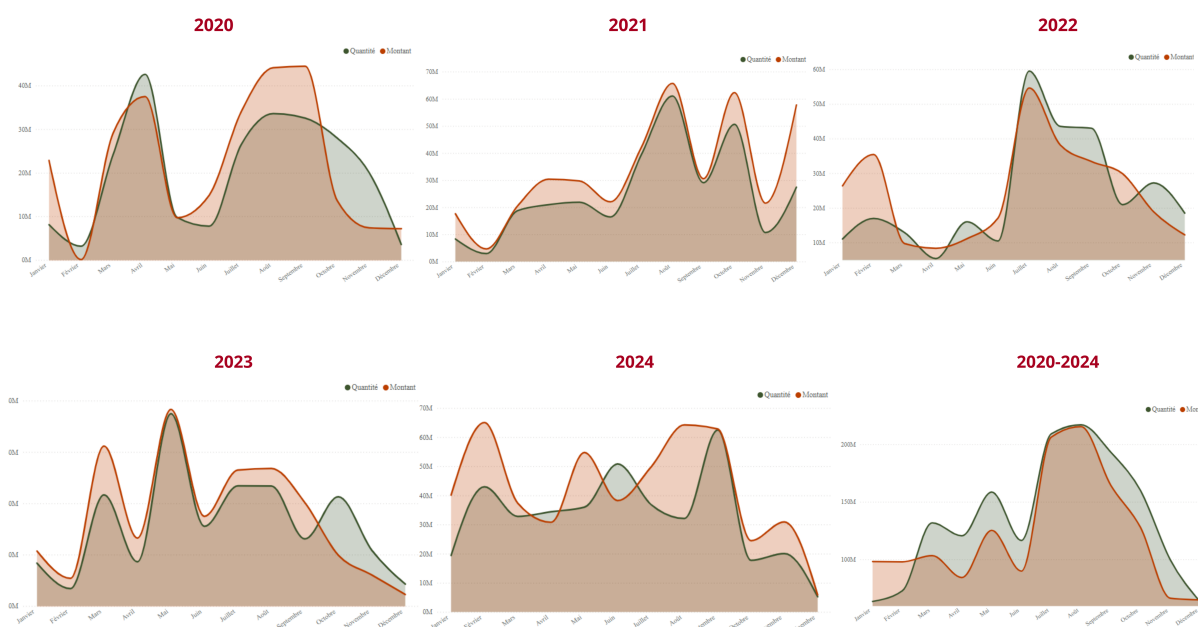


FIGURE 4.16 – Évolution mensuelle de la quantité et du montant des MP achetées (2020–2024).

Sur la période 2020–2024, les mois de juillet à septembre se démarquent par des pics récurrents d’achats, suggérant une saisonnalité ou une anticipation de la production. Une hausse notable des volumes et des montants à partir de 2022 reflète une intensification de l’activité ou une hausse des prix. À l’inverse, les achats sont généralement plus faibles en hiver (décembre à février), traduisant un ralentissement opérationnel. Cette dynamique mensuelle souligne l’importance d’une planification stratégique pour optimiser les approvisionnements et limiter les coûts.

Analyse de la quantité achetée et du montant par MP

Afin d’analyser les volumes et les montants cumulés des achats annuels par matière première, ces données ont été visualisées sous forme d’histogrammes groupés pour chaque année, comme le montre la figure ci-dessous (4.17) :

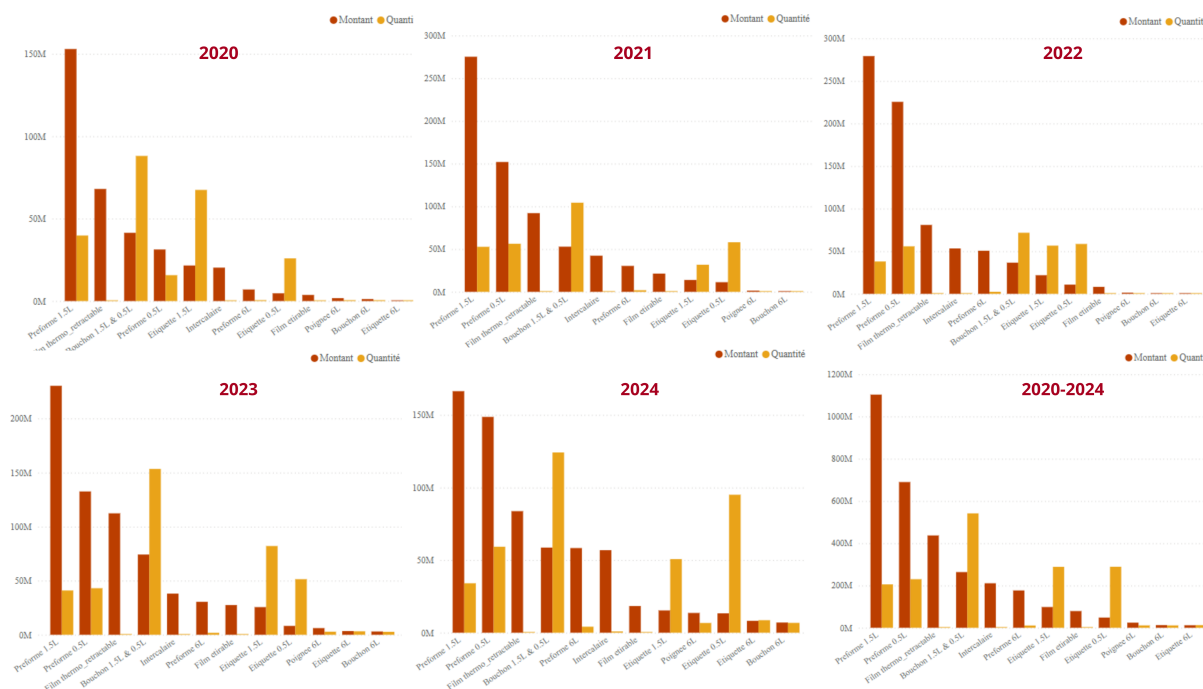


FIGURE 4.17 – Quantité achetée et montant cumulé des MP (2020–2024).

Sur la période 2020–2024, l’analyse des MP révèle une forte concentration des achats autour des formats 0.5L et 1.5L, dominés par les préformes, bouchons et étiquettes, qui représentent à eux seuls la majorité des volumes consommés et des dépenses engagées. Cette standardisation traduit une stratégie de production axée sur les produits les plus courants.

Toutefois, certaines matières comme le film thermo-rétractable ou les préformes 6L affichent un coût unitaire élevé malgré des volumes modestes, signalant des sources potentielles d’optimisation. Ces déséquilibres entre volumes et coûts invitent à une meilleure gestion des achats, notamment en ciblant les matières à fort impact budgétaire.

Globalement, cette structure d’approvisionnement met en lumière la nécessité d’un pilotage affiné des dépenses en fonction du poids stratégique et économique de chaque MP.

Classification des fournisseurs

Afin de mieux structurer la relation fournisseur et de soutenir une stratégie d’approvisionnement ciblée, une segmentation a été réalisée pour chaque MP sur la période 2020–2024, en appliquant l’algorithme de clustering K-means.

Cette approche a permis de hiérarchiser les fournisseurs en fonction de trois critères clés :

- La quantité moyenne achetée.

- Le prix unitaire moyen.
- Le nombre de transactions effectuées.

Trois classes distinctes ont ainsi été identifiées pour chaque MP :

- **Classe A** (fournisseurs prioritaires) : ils fournissent des volumes élevés et/ou présentent une fréquence de commande importante. Leur rôle est essentiel pour assurer la continuité de la production, nécessitant un suivi rigoureux et une gestion proactive.
- **Classe B** (fournisseurs secondaires) : leur contribution est intermédiaire. Ils assurent une flexibilité opérationnelle et complètent l’approvisionnement principal.
- **Classe C** (Fournisseurs alternatifs) : mobilisés de manière ponctuelle ou à faible volume, ils représentent des solutions de repli ou d’appoint.

La classification d’un fournisseur peut varier d’une matière à une autre selon son niveau d’activité sur chaque segment. Cette granularité permet d’adapter les décisions d’achats, les niveaux de suivi et les leviers de négociation en fonction du rôle réel du fournisseur pour chaque MP.

La figure ci-dessous (4.18) présente un aperçu des résultats finaux obtenus par l’algorithme K-means, générés à l’aide de Python :

ID_matiere_premiere	ID_fournisseur	Classe_fournisseur
17	2	1
345	11	35

FIGURE 4.18 – Segmentation des fournisseurs par K-means.

Dans le but de visualiser la distribution des fournisseurs par classe ainsi que les quantités achetées correspondant à chaque segment (A, B et C), deux graphiques en anneau ont été élaborés. La figure ci-dessous (4.19) offre une représentation claire de ces résultats :

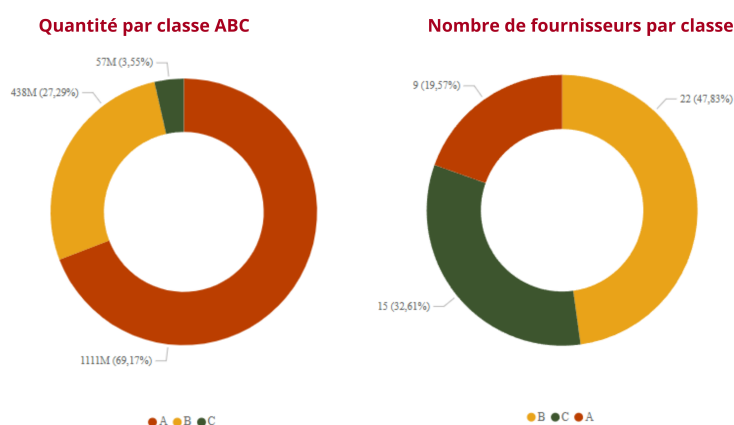


FIGURE 4.19 – Segmentation des fournisseurs et quantités achetées par classe.

Sur la période 2020–2024, l’analyse met en évidence une concentration marquée des volumes d’achat auprès des fournisseurs de classe A, qui représentent à eux seuls 65% des quantités

totales, bien qu'ils ne constituent que 20 % du portefeuille fournisseur. À l'inverse, les classes B et C, comptant respectivement 22 et 15 fournisseurs, couvrent des volumes plus modestes (27 % et 4 %). Cette structuration reflète une forte dépendance à un nombre restreint de partenaires stratégiques, soulignant la nécessité de sécuriser ces relations tout en développant des alternatives pour renforcer la résilience de la chaîne d'approvisionnement.

4.6 Analyse diagnostique

Afin d'identifier les facteurs externes pouvant influencer les dynamiques de vente, une analyse diagnostique a été menée sur la période 2020–2024. Cette démarche vise à croiser les données commerciales avec des variables contextuelles telles que les périodes spéciales (week-ends, jours fériés, Ramadan, événements particuliers), les conditions climatiques (température moyenne et humidité moyenne) ainsi que des variables internes, notamment les écarts entre la production et les ventes.

La figure ci-dessous (4.20) présente un tableau de bord synthétique permettant de visualiser l'impact potentiel de ces facteurs sur les volumes de vente.

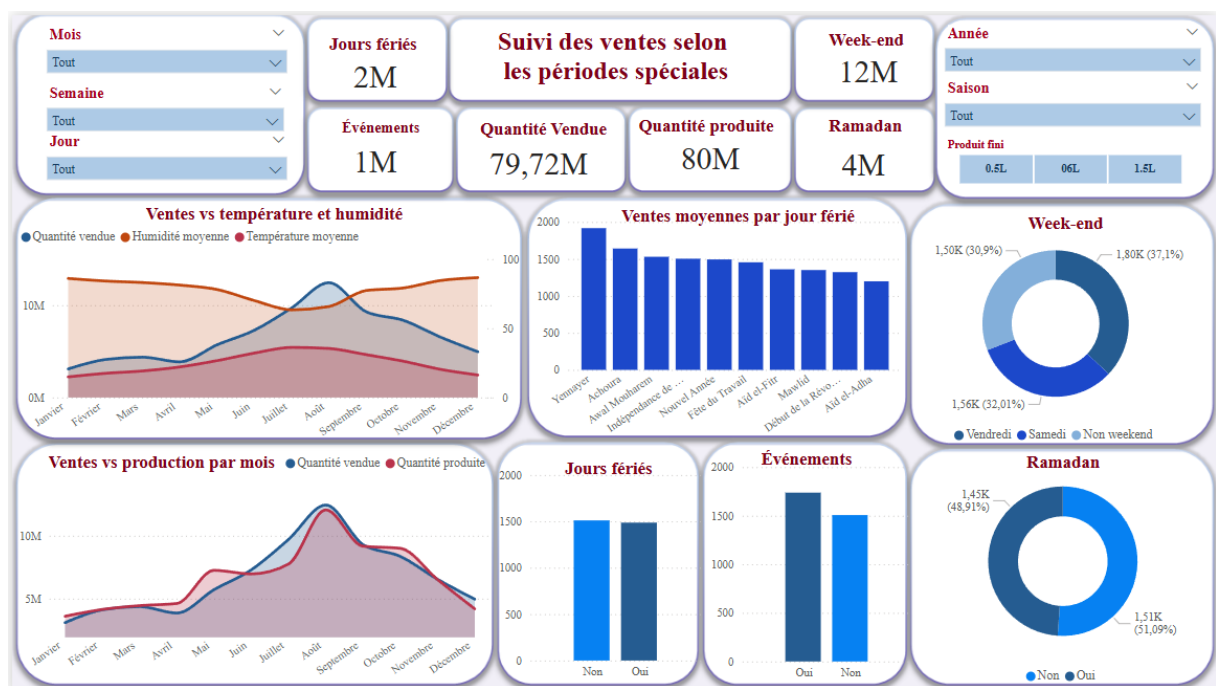


FIGURE 4.20 – Suivi des ventes selon les périodes spéciales et conditions climatiques.

Analyse des ventes vs production

Afin d'évaluer l'alignement entre production et les ventes, une analyse mensuelle a été réalisée sur la période 2020–2024, comparant les volumes produits et vendus. Comme illustré par la figure (4.21) ci-dessous :

Sur la période 2020–2024, les volumes produits et vendus présentent une tendance mensuelle globalement alignée, bien que plusieurs écarts notables soient observés. La production excède

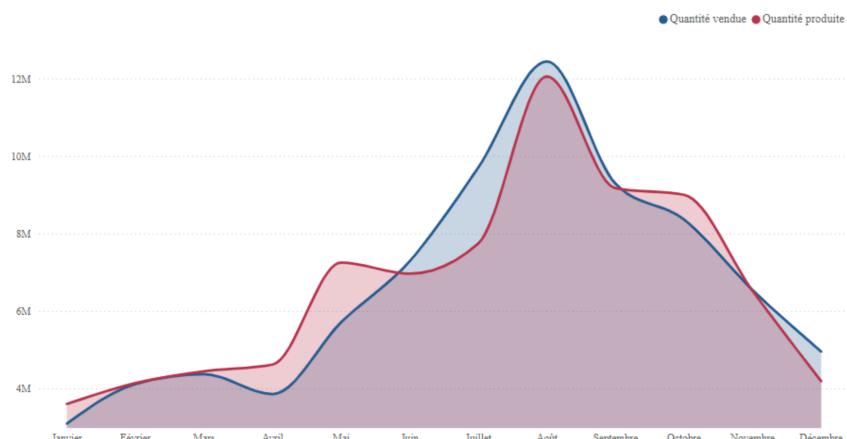


FIGURE 4.21 – Évolution mensuelle des ventes et de la production (2020–2024).

systématiquement les ventes en début d’année (janvier à avril), suggérant une anticipation ou une faible demande saisonnière. À l’inverse, durant les mois estivaux et en décembre, les ventes surpassent la production, indiquant un possible déstockage ou un ajustement insuffisant de la capacité de production. Ces déséquilibres soulignent l’importance d’une meilleure coordination entre production et ventes pour optimiser la gestion des stocks et répondre efficacement à la demande.

Analyse des ventes vs température et humidité

Afin d’examiner l’impact potentiel des conditions climatiques sur les ventes, une analyse croisée des volumes de ventes avec les températures et humidités moyennes mensuelles a été réalisée sur la période 2020–2024, comme illustré par la figure (4.22) ci-dessous :

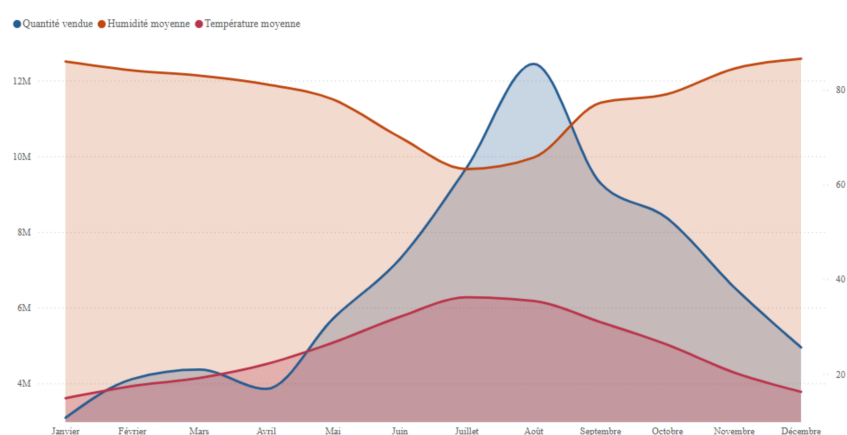


FIGURE 4.22 – Évolution des ventes en fonction de la température et de l’humidité moyennes (2020–2024).

Les résultats révèlent une tendance marquée entre les variables climatiques et les ventes d’eau. La quantité vendue augmente nettement à partir du mois de mai, atteignant un pic en août, en

corrélation avec la hausse des températures (jusqu’à 36,2°C en juillet) et la baisse progressive de l’humidité (jusqu’à 63,3 %). Cette dynamique confirme l’intensification de la demande en période estivale, caractérisée par des conditions chaudes et sèches.

À l’inverse, les mois d’hiver (janvier à mars, puis décembre) enregistrent des niveaux de vente plus faibles, coïncidant avec des températures modérées (autour de 15 °C) et une humidité plus élevée (supérieure à 84 %), ce qui peut freiner la consommation.

Globalement, la température joue un rôle déterminant dans l’évolution des ventes, avec un effet amplifié lorsque l’humidité est faible. Cette analyse confirme l’importance d’intégrer les paramètres météorologiques dans les prévisions de demande afin d’ajuster la production et les stratégies commerciales de manière proactive.

Analyse des ventes selon les périodes spéciales

Afin de mieux comprendre l’influence du calendrier sur la consommation d’eau, une analyse des ventes moyennes a été réalisée en fonction de plusieurs facteurs temporels spécifiques (les jours fériés nationaux et religieux, les événements particuliers, la période du Ramadan, ainsi que les week-ends) comme le montre la figure ci desous :

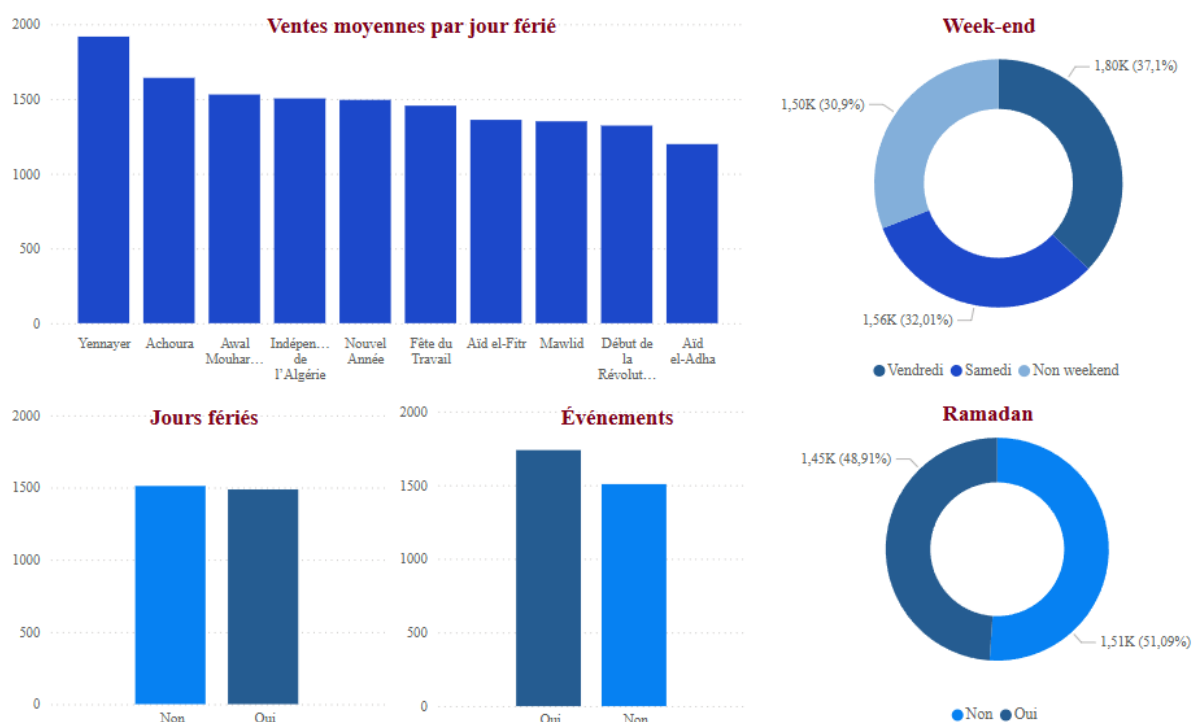


FIGURE 4.23 – Évolution des ventes moyennes selon les périodes spéciales (2020–2024).

L’analyse révèle que les ventes varient sensiblement selon les jours du calendrier. En moyenne, les jours fériés enregistrent une légère baisse des ventes (1 486 unités) par rapport aux jours non fériés (1 511 unités). Toutefois, certains jours comme Yennayer (1 917) et Achoura (1 641) génèrent des pics de consommation, tandis que les fêtes religieuses comme Aïd el-Adha (1 198) affichent des niveaux plus bas.

Les événements spéciaux stimulent clairement la demande, avec une moyenne de 1 738 unités vendues contre 1 507 en temps normal. De même, les week-ends, en particulier le vendredi (1 803), se distinguent par des volumes élevés, probablement en raison des achats anticipés par les grossistes.

Pendant le Ramadan, une baisse de la consommation est observée (1 449 unités), ce qui peut être attribué aux changements de rythme alimentaire et à la réduction des périodes d'hydratation.

Il est important de noter qu'Ovitalé vend exclusivement aux grossistes, qui anticipent souvent la demande finale. Ainsi, certains pics de vente peuvent précéder les événements réels, reflétant davantage les stratégies d'approvisionnement des intermédiaires que le comportement direct des consommateurs finaux.

En conclusion, ces résultats soulignent l'intérêt d'intégrer les paramètres calendaires dans les modèles de prévision pour mieux planifier la production, ajuster les stocks et optimiser les actions commerciales.

4.7 Analyse prédictive

À la suite de l'analyse descriptive qui a permis de mieux comprendre les tendances passées des ventes, de la production et des niveaux de stock, cette section se concentre sur une approche prédictive visant à anticiper la demande pour le premier trimestre 2025.

En s'appuyant sur des modèles de prévision basés sur le ML, cette approche permet d'anticiper l'évolution de la demande à court terme, afin de mieux comprendre les dynamiques futures et d'orienter les décisions de gestion de manière proactive.

4.7.1 Préparation des données

À l'issue des phases de collecte et de traitement des données de ventes, un dataset de référence a été constitué, comprenant 12 colonnes et 4104 lignes :

- **date** : date de la transaction de vente.
- **article** : identifiant du format de PF (1003 : 0.5 L, 1002 : 1.5 L, 1006 : 6 L).
- **quantite** : quantité vendue exprimée en nombre de fardeaux.
- **prix_unitaire** : prix de vente unitaire de l'article.
- **nombre_clients** : nombre total de clients dans la wilaya concernée.
- **temperature_max** : température maximale moyenne enregistrée le jour de la vente.
- **humidite_max** : humidité maximale moyenne enregistrée le jour de la vente.
- **jour_ferie** : indique si la date correspond à un jour férié (*Oui = 1/Non = 0*).
- **weekend** : indique si la date correspond à un week-end (*Oui = 1/Non = 0*).
- **ramadan** : indique si la date se situe pendant le mois de Ramadan (*Oui = 1/Non = 0*).
- **evenement** : présence d'un événement particulier ce jour-là (*Oui = 1/Non = 0*).
- **saison** : saison correspondant à la date (hiver, printemps, été, automne).

	date	article	quantite	prix_unitaire	nombre_clients	temperature_max	humidite_max	jour_ferie	weekend	ramadan	evenement	saison
0	2020-01-02	1002	30016	92	24	15	86	0	0	0	0	Hiver
1	2020-01-02	1003	2688	115	12	15	87	0	0	0	0	Hiver

FIGURE 4.24 – Jeu de données de ventes.

Une phase de préparation a ensuite été réalisée afin de rendre ces données exploitables dans le cadre de l’analyse prédictive. Elle a comporté les étapes suivantes :

- Encodage de la variable catégorielle *saison* à l’aide de la méthode *One-Hot Encoding*.
- Standardisation des variables numériques *prix_unitaire*, *nombre_clients*, *temperature_max* et *humidite_max*.

Le dataset a ensuite été segmenté en trois sous-ensembles distincts, chacun correspondant à un article spécifique :

- Un dataset pour l’article *1002*, de taille (1493, 13)
- Un dataset pour l’article *1003*, de taille (1448, 13)
- Un dataset pour l’article *1006*, de taille (1163, 13)

	quantite	prix_unitaire	nombre_clients	temperature_max	humidite_max	jour_ferie	weekend	ramadan	evenement	automne	hiver	printemps	ete
0	30016	-1.168220	-0.192409	-1.319135	0.312690	0	0	0	0	0	1	0	0
1	22848	-1.168449	-0.648751	-1.674182	0.891250	0	1	0	0	0	1	0	0

FIGURE 4.25 – Exemple d’un jeu de données d’entraînement de l’article *1002*.

La sélection des variables explicatives s’est appuyée sur une double démarche, une collaboration avec les experts métiers d’Ovitalé, ainsi qu’une analyse diagnostique des données. Cette démarche conjointe a permis d’identifier 12 variables explicatives, à la fois internes et contextuelles, exerçant une influence significative sur la demande, ainsi qu’une variable cible, la *quantité*, que l’on cherche à prédire.

4.7.2 Séparation des données

Pour garantir une évaluation fiable des modèles de prévision, chaque jeu de données, correspondant à un article spécifique, a été divisé en deux sous-ensembles distincts :

- **Ensemble d’entraînement (80%)** : utilisé pour entraîner les modèles et modéliser les relations entre les variables explicatives et la variable cible.
- **Ensemble de test (20%)** : réservé à la validation finale des performances prédictives sur des données inédites.

4.7.3 Choix et évaluation des modèles de prévision

Pour anticiper la demande hebdomadaire sur la période de janvier à mars 2025, différentes familles de modèles ont été explorées afin d’identifier la solution la plus performante et la plus adaptée à la nature des données.

Une première série de tests a été réalisée à l’aide de modèles statistiques classiques souvent utilisés pour les séries temporelles :

- Autoregressive Integrated Moving Average (ARIMA).
- Seasonal Autoregressive Integrated Moving Average (SARIMA).
- Holt-Winters.
- Facebook Prophet.

Bien que ces modèles soient conçus pour modéliser efficacement les tendances et les effets saisonniers, leurs performances se sont avérées limitées sur le jeu de données disponible. Les résultats obtenus à travers les métriques d'évaluation (RMSE, MAE, MSE et R^2) ont révélé une précision insuffisante. Cette contre-performance s'explique principalement par des problèmes de qualité des données, notamment la taille réduite de l'échantillon d'entraînement et l'incomplétude de la série temporelle, marquée par l'absence de plusieurs jours de vente. Ces lacunes compromettent la continuité et la représentativité des données, éléments essentiels pour les modèles de prévision.

Face aux limites des modèles statistiques, l'analyse a été étendue à un ensemble de modèles de ML supervisés capables de mieux exploiter les relations complexes entre les variables explicatives et la variable cible :

- **Modèles linéaires** : Régression Linéaire, Ridge, Lasso.
- **Modèles non linéaires** : Support Vector Regression (SVR), K-Nearest Neighbors (KNN), Arbres de décision.
- **Modèles d'ensemble** : Random Forest, Gradient Boosting, Extreme Gradient Boosting (XGBoost), Light Gradient Boosting Machine (LightGBM), Adaptive Boosting (AdaBoost), Bootstrap Aggregating (Bagging).
- **Réseaux de neurones** : Multi-Layer Perceptron Regressor (MLPRegressor).

Une méthodologie rigoureuse a été appliquée à chaque modèle afin d'assurer des résultats fiables :

- **Validation croisée à 5 plis** : pour éviter le surapprentissage et évaluer la stabilité des performances.
- **Optimisation des hyperparamètres via *GridSearchCV*** : une recherche systématique a été conduite pour chaque algorithme afin d'identifier la combinaison optimale de paramètres.

Les modèles ont été comparés à l'aide des métriques standards de régression, notamment la RMSE, la MAE et le R^2 , afin d'identifier ceux offrant les meilleures performances prédictives.

4.7.4 Résultats obtenus

Pour chaque article nous avons sélectionné un meilleur modèle ayant obtenu les meilleures performances selon les métriques d'évaluation standard.

Article 1002

Le modèle Gradient Boosting se démarque pour l'article *1002*, en atteignant un R^2 de (0.7871), indiquant une capacité élevée à modéliser la dynamique des données. Sa capacité à capturer des relations non linéaires complexes entre les variables explicatives en fait un choix optimal.

En comparaison, les modèles linéaires et le SVR ont montré une sous-performance notable, ne parvenant pas à restituer la variabilité structurelle des séries.

Les performances de certains modèles testés sont synthétisées dans le tableau (4.7) ci-dessous :

Modèle	MAE	MSE	RMSE	R^2
Gradient Boosting	6713.01	112M	10 595	0.7871
Bagging	6740.37	117M	10 828	0.7777
SVR	8833.10	175M	13 247	0.6675
Ridge	8794.39	179M	13 391	0.6603

TABLE 4.7 – Performances des modèles de ML pour l'article 1002.

En complément, la figure (4.26) illustre la qualité des prédictions du modèle Gradient Boosting :

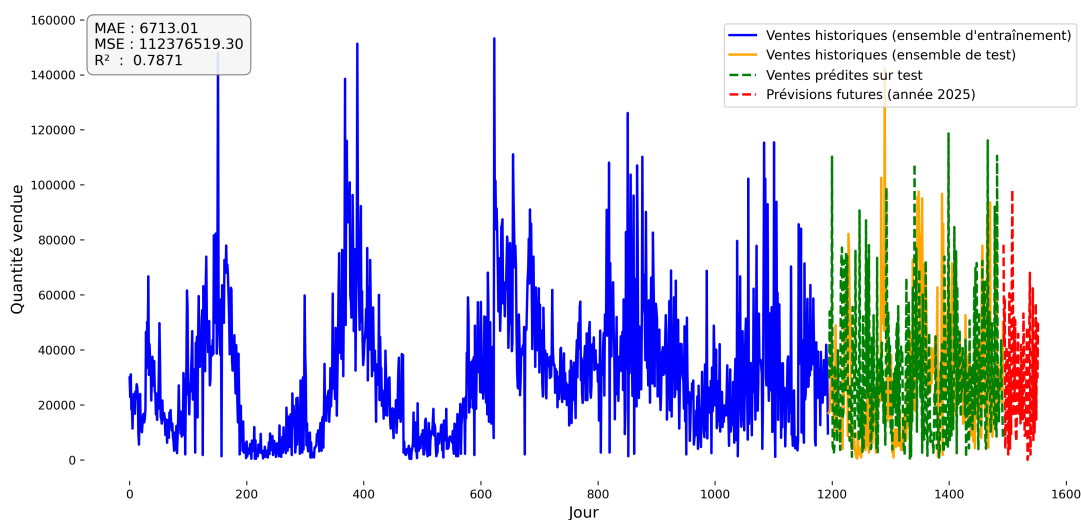


FIGURE 4.26 – Prévisions de la demande pour l'article 1002

L'analyse historique de l'article 1002 met en évidence une forte saisonnalité, caractérisée par des pics récurrents aux mêmes périodes chaque année. Le modèle développé parvient à capturer efficacement cette dynamique, en expliquant près de 79 % de la variance des ventes et en affichant des erreurs de prévision modérées (MAE = 6 713, RMSE = 10 595). Bien que les résultats soient globalement satisfaisants, certaines fluctuations irrégulières demeurent inexplicables, probablement en raison de la qualité imparfaite et de la taille limitée des données utilisées. L'intégration de données plus riches, variées et fiables permettrait d'améliorer la précision des prévisions. En somme, les prévisions relatives à l'article 1002 pour le premier trimestre 2025 s'alignent avec les tendances historiques et constituent une base solide pour appuyer la prise de décision.

Article 1003

Le modèle MLPRegressor affiche les meilleures performances pour l'article 1003, avec un R^2 de (0.7490). Sa capacité d'apprentissage profond lui permet de modéliser efficacement des relations non linéaires complexes. À l'inverse, les modèles plus simples comme le KNN ou l'arbre de décision ont montré des limites de généralisation, affectant leur précision.

Les performances de certains modèles testés sont synthétisées dans le tableau (4.8) ci-dessous :

Modèle	MAE	MSE	RMSE	R^2
MLPRegressor	4729.08	84.7M	9201.91	0.7490
Bagging Regression	4656.83	91.2M	9552.29	0.7296
KNN	5713.98	138.2M	11 754.12	0.5906
Decision Tree	5452.19	138.4M	11 765.27	0.5898

TABLE 4.8 – Performances des modèles de ML pour l'article 1003.

En complément, la figure (4.27) illustre la qualité des prédictions du modèle Gradient Boosting :

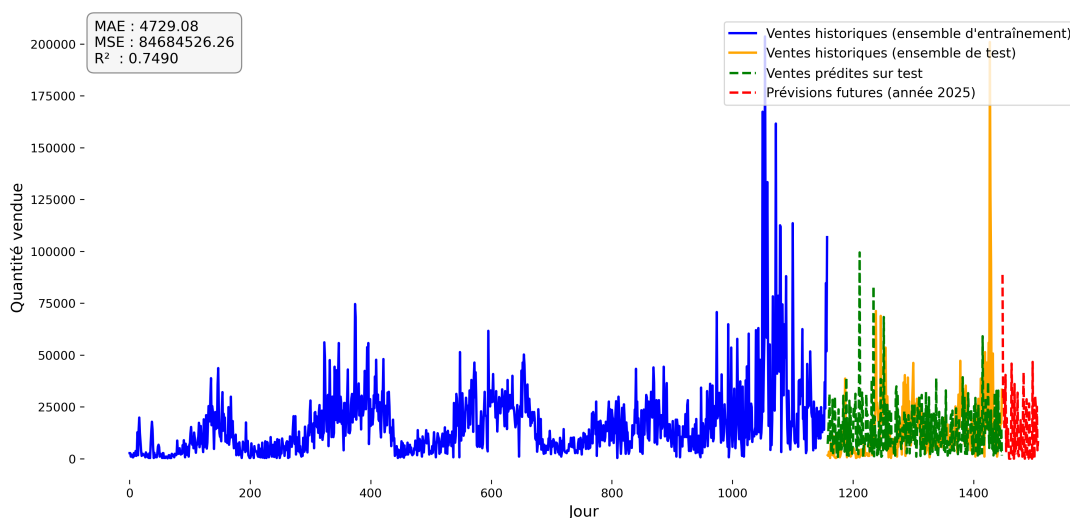


FIGURE 4.27 – Prévisions de la demande pour l'article 1003

L'analyse des ventes historiques de l'article 1003 révèle une dynamique marquée par des tendances variables et une saisonnalité bien définie. Le modèle développé capture efficacement ces caractéristiques, expliquant environ 75 % de la variance des ventes, avec des erreurs de prévision modérées (MAE = 4 729, RMSE = 9 202). Malgré ces résultats satisfaisants, certaines fluctuations irrégulières restent partiellement inexpliquées, probablement en raison de limitations dans la qualité, la granularité et la taille des données. L'intégration de données plus riches et plus fiables renforcerait la précision des prévisions. Globalement, les estimations pour le premier trimestre 2025 s'alignent avec les tendances passées et constituent un appui pertinent pour la planification à court terme.

Article 1006

Le modèle Bagging s'impose comme le plus performant pour l'article 1006, avec un R^2 de (0.8206). Sa robustesse face aux variations structurelles du signal et sa capacité à réduire la variance des prédictions grâce à l'agrégation d'estimateurs faibles en font un excellent choix. Il surpasse même XGBoost et Random Forest sur cet article.

Les performances de certains modèles testés sont synthétisées dans le tableau (4.9) ci-dessous :

Modèle	MAE	MSE	RMSE	R^2
Bagging Regression	2792.29	25.5M	5044.86	0.8206
XGBoost	2678.81	25.6M	5057.43	0.8197
Random Forest	2840.27	27.2M	5214.25	0.8084
Gradient Boosting	2832.94	28.7M	5360.30	0.7975

TABLE 4.9 – Performances des modèles pour l'article 1006

En complément, la figure (4.28) illustre la qualité des prédictions du modèle Gradient Boosting :

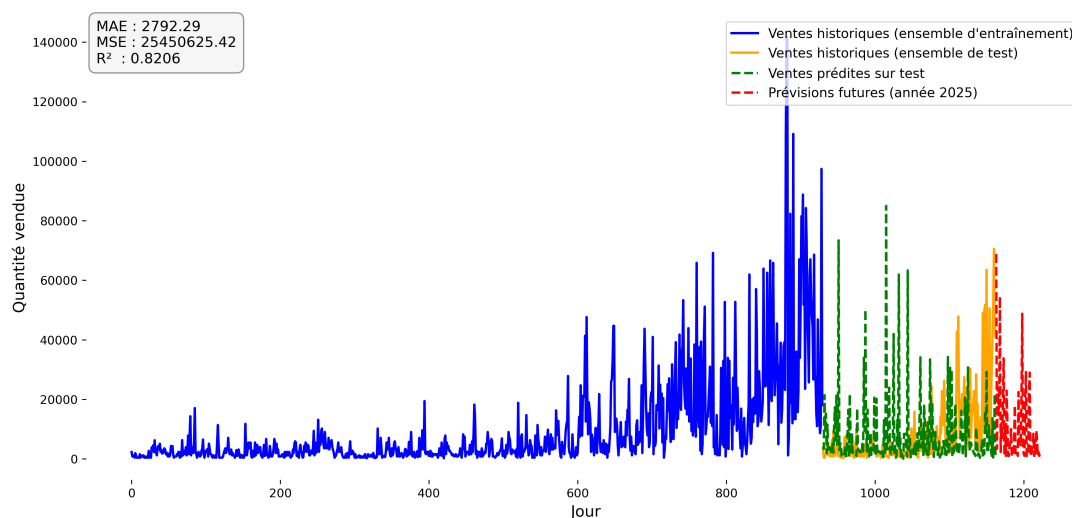


FIGURE 4.28 – Prévisions de la demande pour l'article 1003

Les prévisions de ventes pour le premier trimestre 2025 s'alignent fidèlement avec les tendances historiques, marquées par une progression notable en janvier, suivie d'une stabilisation en février et mars. Le modèle parvient à expliquer 82 % de la variance des ventes, témoignant d'une bonne performance globale, bien que 18 % des fluctuations demeurent inexpliquées, principalement en raison de la qualité imparfaite et du volume limité des données disponibles. L'enrichissement du jeu de données, tant en diversité qu'en fiabilité, constituerait une piste d'amélioration pertinente. En somme, ces prévisions constituent une base solide pour la planification à court terme, tout en offrant des perspectives d'optimisation méthodologique.

4.7.5 Visualisation interactive des prévisions

Afin de faciliter l’exploitation des résultats par les décideurs, un tableau de bord interactif (4.29) a été conçu. Celui-ci permet de visualiser les prévisions de la demande de manière claire et dynamique. Grâce à des filtres personnalisables, il offre une vue détaillée et flexible des ventes projetées, permettant ainsi une prise de décision plus rapide et mieux informée.



FIGURE 4.29 – Suivi de la demande future.

L’analyse prédictive montre une demande mensuelle stable, avec un pic en janvier, un repli en février, suivi d’une reprise en mars. Le format 1,5 L domine largement le marché (56% de la demande), affirmant son rôle central dans la stratégie de production. Le 0,5 L reste un format secondaire régulier, tandis que le 6 L affiche une demande plus faible.

Sur le plan hebdomadaire, d’importantes variations sont observées, notamment des pics en semaines 2, 4, 5 et 8, et des creux autour des semaines 6, 7 et 11. Ces dynamiques permettent de mieux ajuster la planification de la production et des approvisionnements en fonction des formats et des périodes.

4.8 Analyse prescriptive

À la suite de l’analyse prédictive, qui a permis d’anticiper la demande hebdomadaire du premier trimestre 2025, cette section d’analyse prescriptive vise à transformer les prévisions obtenues en actions concrètes à forte valeur opérationnelle. Elle s’articule autour de plusieurs axes clés, l’élaboration d’un plan de production optimal, la planification de l’approvisionnement et la gestion des stocks, ainsi que la recommandation des fournisseurs. En intégrant les contraintes industrielles et logistiques, les coûts associés et les classifications stratégiques des fournisseurs, cette approche permet de soutenir des prises de décision éclairées.

L'ensemble est appuyé par des visualisations interactives facilitant l'exploration des résultats et le pilotage global de la chaîne d'approvisionnement.

4.8.1 Planification de la production

En s'appuyant sur le modèle du lot économique de production (EPQ) un PDP hebdomadaire optimisé a été élaboré. L'objectif est de minimiser les coûts totaux de fabrication et de stockage, tout en garantissant la satisfaction de la demande anticipée.

Ce modèle tient compte simultanément du rythme de production et du rythme de consommation, ce qui permet de déterminer la quantité optimale à produire par cycle, tout en limitant les coûts de lancement de production et de stockage intermédiaire.

Les différentes formules mobilisées sont détaillées ci-dessous :

Quantité optimale à produire

Quantité optimale à fabriquer lors de chaque cycle de production, afin de satisfaire la demande tout en minimisant les coûts :

$$Q^* = \sqrt{\frac{2 \cdot L \cdot D}{C \cdot \left(\frac{p-u}{p}\right)}}$$

Nombre de cycles de production

Nombre total de cycles nécessaires sur la période pour couvrir la demande anticipée :

$$N = \frac{D}{Q^*}.$$

Quantité maximale en stock

Niveau maximal de stock atteint durant un cycle :

$$Q_{max} = \left(\frac{p-u}{p}\right) \cdot Q^*$$

Durée de production (heure)

Temps nécessaire pour produire une quantité Q^* :

$$Durée_{prod} = \frac{Q^*}{p} \cdot 168.$$

Durée de consommation (heure)

Période pendant laquelle le stock produit sera écoulé selon le rythme de la demande :

$$Durée_{cons} = \frac{Q_{max}}{u} \cdot 168.$$

Coût total optimal (DA)

Estimation du coût total par cycle, intégrant les frais de stockage et les frais de lancement de production :

$$Coût_{total} = C \cdot \left(\frac{Q_{max}}{2}\right) + L \cdot N,$$

où :

- D : Demande hebdomadaires (estimée à partir des prévisions hebdomadaires).
- p : Taux de production (cadence de ligne de production).
- u : Taux de consommation (la demande hebdomadaire).
- L : Coût de lancement d'un cycle de production.
- C : Coût unitaire de stockage d'une bouteille par semaine.

Afin de concrétiser le modèle théorique et de le traduire en décisions opérationnelles, un script Python a été développé pour appliquer les formules du lot économique de production à chaque format de PF. La figure ci-dessous (4.30) présente un aperçu final du dataset généré :

	ID_article	ID_date	demande	Q_optimal	Nb_cycles	Q_max	Durée_Prod	Durée_Cons	Cout_Total_Optimal
0	1002	5	469362	142750	3	117428	9	42	658
1	1002	12	988362	237372	4	148706	15	25	832

FIGURE 4.30 – Aperçu hebdomadaire du plan de production optimal pour l'article 1002.

Le tableau ci-dessous présente une synthèse du plan de production optimal hebdomadaire pour les trois formats de PF :

Format	Demande	N	$Durée_{prod}$	$Durée_{cons}$	$Coût_{total}$	Production
1.5 L	10 979 664	51	174	383	10 248	10 976 009
0.5 L	9 519 788	37	436	335	7 414	9 520 359
06 L	640 106	12	374	2 125	2 510	639 849

TABLE 4.10 – Synthèse des résultats d'optimisation pour les différents formats de production

L'analyse globale des trois formats de produits finis révèle des dynamiques de consommation et de production distinctes. Les formats 1,5 L et 0,5 L se caractérisent par une demande élevée et une rotation rapide, nécessitant une production régulière et soutenue pour répondre efficacement au rythme de consommation. Le format 1,5 L, grâce à des durées de production courtes et des volumes importants, affiche une excellente synchronisation entre production et écoulement. Le format 0,5 L, malgré une cadence machine plus faible allongeant le temps de production, évite le surstock grâce à une absorption rapide par le marché.

En revanche, le format 6 L, marqué par une demande plus modérée et une consommation étalée dans le temps, présente un risque plus élevé de surstock. Cela implique une stratégie de production plus prudente, visant à limiter les invendus et à optimiser l'espace de stockage.

Le plan de production respecte les capacités des lignes sans dépasser les contraintes, maîtrise les coûts et optimise les cycles pour limiter les pertes. Il met en avant une gestion adaptée, accélérer la production des formats à forte demande et ajuster précisément celle des formats moins consommés pour éviter les surplus.

4.8.2 Visualisation interactive de la planification de la production

Afin de faciliter l’exploitation des résultats par les décideurs, un tableau de bord interactif (4.31) a été conçu. Celui-ci permet de visualiser le plan de production de manière claire et dynamique.

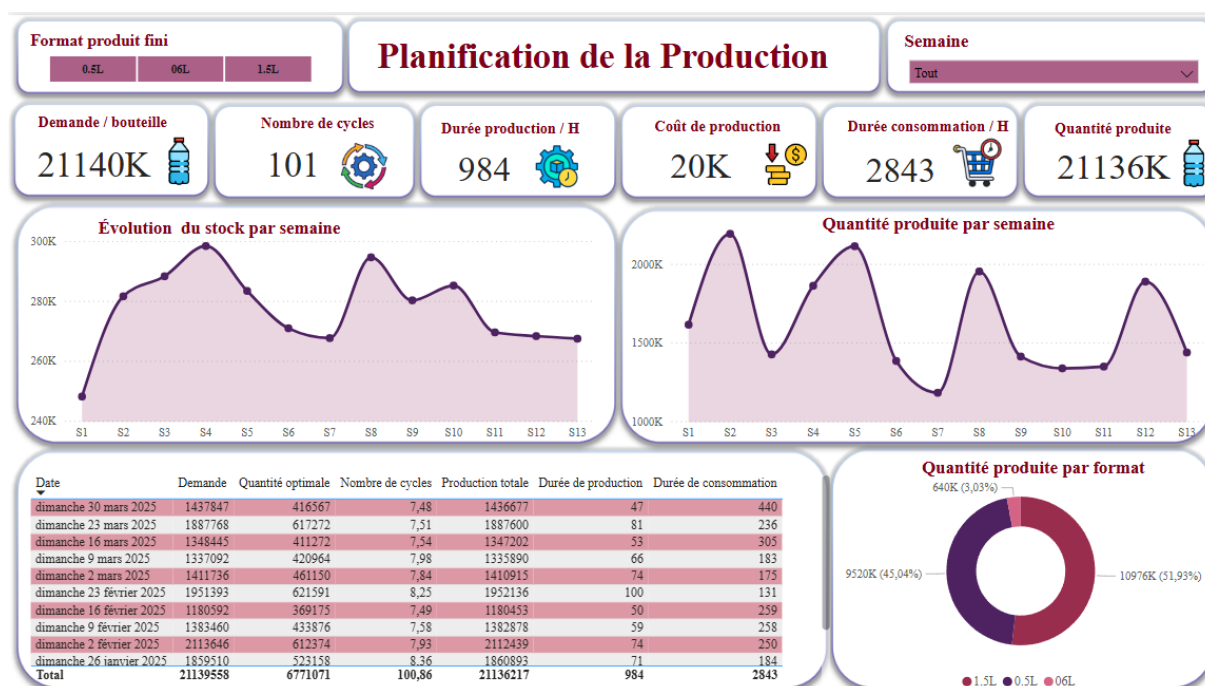


FIGURE 4.31 – Planification de la production.

4.8.3 Planification de l’approvisionnement et gestion des stocks

À partir du plan de production hebdomadaire établi, une planification des besoins en MP a été réalisée selon la méthode MRP. Cette démarche vise à assurer la disponibilité des MP en temps voulu, tout en limitant les surstocks et en évitant les ruptures, afin d’optimiser à la fois les coûts logistiques et la fluidité du processus de production. La méthode MRP permet de déterminer à la fois la quantité optimale à commander et le moment opportun pour lancer chaque commande, en prenant en compte les éléments suivants :

- Les besoins prévisionnels issus du PDP.
- Le Stock Initial (SI) disponible pour chaque MP.
- Le délai d’approvisionnement (fixé ici à deux semaines).
- Le Stock de Sécurité (SS), défini pour pallier toute incertitude ou retard.

Les différentes formules utilisées dans cette démarche sont présentées ci-dessous :

EOQ

La EOQ correspond au volume de commande qui permet de minimiser le coût total d’approvisionnement, en équilibrant les coûts de passation des commandes et les coûts de stockage. Elle est calculée selon la formule suivante :

$$Q^* = \sqrt{\frac{2 \times L \times D}{C}}$$

où :

- *L* : coût de passation d’une commande.
- *C* : coût unitaire de stockage.

Détermination des Besoins Bruts (BB)

Avant de procéder au calcul des BB, une nomenclature détaillée a été établie pour chaque PF, comme illustré dans la figure ci-dessous :

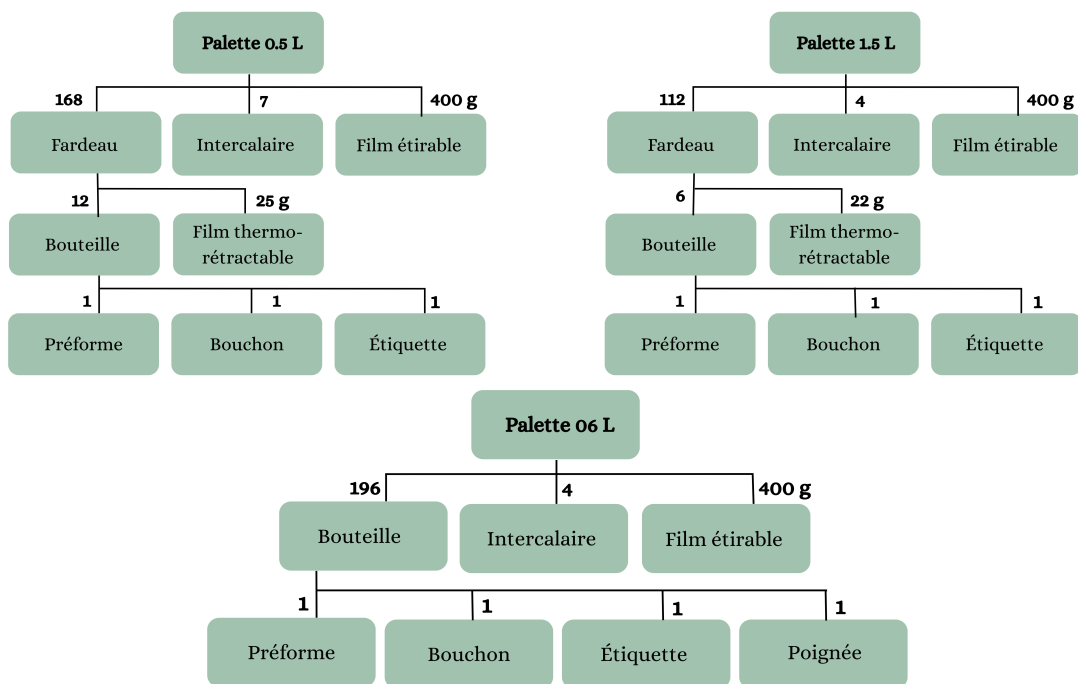


FIGURE 4.32 – Nomenclature des PF.

À partir des résultats obtenus via le PDP, les BB ont été calculés pour chaque MP en appliquant la formule suivante :

$$BB = Q^* \times N \times Quantité,$$

où :

- *Quantité* : quantité de MP nécessaire pour produire une unité de PF.

Stock Prévisionnel (SP)

Le Stock Prévisionnel (SP) permet d'estimer le niveau de stock disponible à la fin de chaque semaine. Il est déterminé selon la formule suivante :

$$SP = Stock_initial + Réceptions - BB.$$

Besoins Nets (BN)

Le Besoins Nets (BN) représente la quantité manquante de MP pour satisfaire les besoins, après prise en compte du stock disponible. Il est calculé comme suit :

$$BN = SS - SP.$$

Quantité à commander

La quantité à commander correspond au volume à approvisionner afin de couvrir le BN, tout en respectant le EOQ. Elle est calculée selon la formule suivante :

$$Quantité_commander = \left\lceil \frac{BN}{Q^*} \right\rceil \times Q^*.$$

La commande est planifiée de manière à tenir compte du délai d'approvisionnement, afin d'assurer la disponibilité des MP au moment requis.

Coûts d'approvisionnement

Les résultats du MRP permet d'estimer les coûts d'approvisionnement, en se basant sur les indicateurs suivants :

— **Coût total de commande :**

$$Coût_c = N_c \times L.$$

— **Coût total de stockage :**

$$Coût_s = C \times S_m.$$

— **Coût total :**

$$Coût_t = Coût_c + Coût_s.$$

où :

- N_c : le nombre de commandes.
- L : le coût unitaire de passation.
- C : le coût unitaire de stockage.
- S_m : le stock moyen sur la période.

Dans le but de traduire le modèle théorique en décisions opérationnelles concrètes, un script Python a été développé, permettant de planifier efficacement les besoins en MP et d'estimer les coûts globaux d'approvisionnement. Les figures ci-dessous présente un aperçu des jeux de données finaux générés à l'issue de cette automatisation :

ID_date	ID_matiere_premiere	BB	SP	BN	Quantite_commander	Quantite_recue
0	0	1	0	46120703	0	0
1	5	1	1535118	44585585	0	0

FIGURE 4.33 – Aperçu hebdomadaire de la planification MRP des MP.

ID_matiere_premiere	Nombre de commandes	Coût de commande	Coût de stockage	Coût total
0	1	0	0	497877
1	2	1	100	10024

FIGURE 4.34 – Aperçu des coûts d’approvisionnement des MP.

Le tableau ci-dessous (4.11) présente une synthèse de la planification de l’approvisionnement et de la gestion des stocks en MP :

MP	BB	BN	SI	SS	Quantité commandée	Coût _t
Intercalaire	802K	1.14M	361K	828K	849K	18K
Preforme 1.5L	11.0M	1.00M	13.9M	5.29M	2.10M	124K
Bouchon 6L	640K	721K	309K	529K	506K	10K
Poignée 6L	640K	632K	354K	529K	506K	11K
Film étirable	63K	1K	81K	68K	79K	2K
Bouchon 1.5L & 0.5L	20.5M	0	46.1M	15.9M	0	498K
Étiquette 0.5L	9.52M	0	18.1M	2.65M	0	183K
Étiquette 1.5L	11.0M	0	24.3M	5.29M	0	262K
Étiquette 6L	640K	0	2.50M	529K	0	30K
Film thermo_rétractable	24K	0	50K	19K	0	519
Preforme 0.5L	9.52M	0	16.8M	2.65M	0	165K
Preforme 6L	640K	0	1.46M	529K	0	15K

TABLE 4.11 – Synthèse de la planification de l’approvisionnement et de la gestion des stocks en MP.

Le tableau met en lumière une concentration marquée des BB sur les MP liées aux formats 0,5L et 1,5L notamment les préformes, bouchons et étiquettes ce qui reflète leur forte demande et leur rotation rapide. À l’inverse, les composants dédiés au format 6L, tels que les bouchons et les poignées, présentent des besoins plus modérés, traduisant une consommation plus lente.

L'analyse des BN met en évidence la nécessité d'un réapprovisionnement proactif pour certaines références sensibles, notamment l'intercalaire et les préformes 1,5L, afin de prévenir toute rupture. De manière générale, les stocks initiaux demeurent au-dessus des seuils de sécurité, assurant ainsi une marge de manœuvre face aux aléas de la demande ou aux retards d'approvisionnement.

Les quantités à commander sont ajustées de manière précise, permettant de garantir la continuité de la production tout en évitant les surstocks. Cette approche méthodique favorise un équilibre optimal entre fluidité des opérations, maîtrise des coûts et efficacité des niveaux de stock, contribuant à l'amélioration globale de la performance logistique.

4.8.4 Visualisation interactive de la planification d'approvisionnement et de la gestion des stocks

Dans le but de faciliter l'analyse et l'interprétation des résultats par les décideurs, un tableau de bord interactif a été conçu. Il permet une visualisation dynamique et synthétique de la planification d'approvisionnement et de la gestion des stocks.

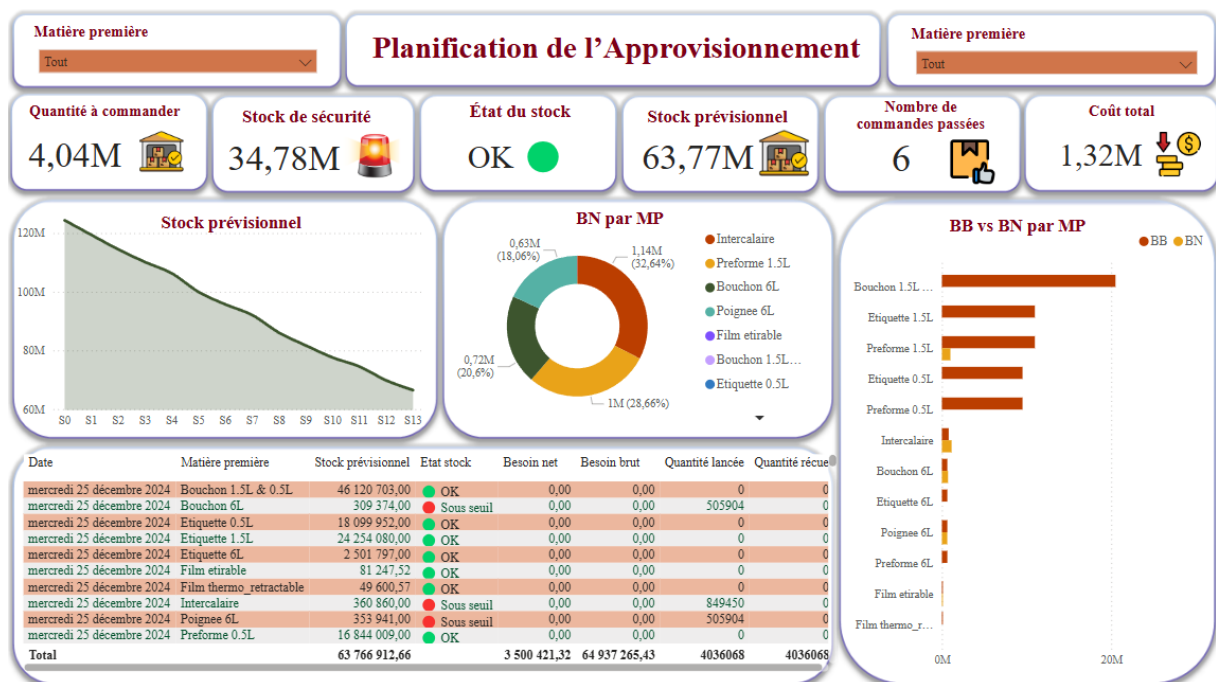


FIGURE 4.35 – Planification de l'Approvisionnement.

4.8.5 Recommandation des fournisseurs

À partir de la segmentation des fournisseurs réalisée dans l'analyse descriptive, ceux-ci ont été classés en trois catégories stratégiques (A, B et C). En parallèle, sur la base des quantités à commander calculées via le MRP, une stratégie de répartition des commandes a été élaborée. Les volumes à commander ont été distribués entre les fournisseurs selon un pourcentage défini en fonction de leur classe :

- **Classe A** : 70 % de la quantité totale à commander.
- **Classe B** : 20 % de la quantité totale à commander.
- **Classe C** : 10 % de la quantité totale à commander.

La figure ci-dessous (4.36) présente un aperçu des résultats générés : pour chaque MP, le ou les fournisseurs recommandés ainsi que les quantités attribuées, calculées automatiquement à l'aide de Python.

ID_date	ID_matiere_premiere	ID_fournisseur	Classe_fournisseur	Quantite_recommandee	
545	5	6	21	B	8366.526316
546	5	6	22	C	4183.263158

FIGURE 4.36 – Recommandation des fournisseurs et répartition des quantités par MP.

Cette méthode permet d’assurer une priorité aux partenaires stratégiques, tout en maintenant une certaine diversification des sources d’approvisionnement pour limiter les risques.

4.8.6 Visualisation interactive des recommandations de fournisseurs

Dans le but de faciliter la prise de décision et d’optimiser la gestion des approvisionnements, un tableau de bord interactif a été conçu (4.37) afin de visualiser de manière claire et dynamique les recommandations de fournisseurs. Cette interface permet aux décideurs d’identifier rapidement, pour chaque MP, les fournisseurs à privilégier ainsi que les volumes à leur attribuer, en tenant compte de leur classification stratégique (classes A, B et C).

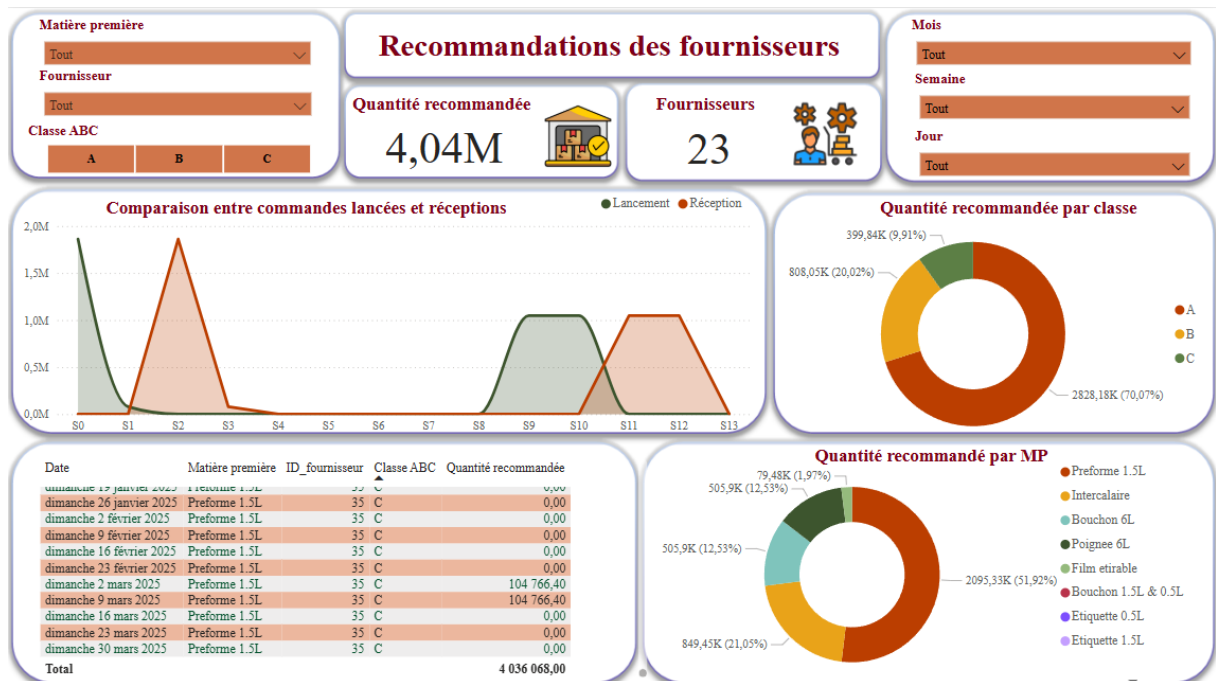


FIGURE 4.37 – Recommandation des fournisseurs.

4.9 Pipeline Analytique : De Python à Power BI

Dans le cadre de cette étude, Python a été utilisé comme outil principal pour le traitement des données et la génération des résultats. L'ensemble des algorithmes de ML ainsi que ceux liés à l'optimisation de la chaîne logistique ont été entièrement développés en Python, en s'appuyant sur diverses bibliothèques spécialisées.

Les résultats obtenus à chaque étape du processus ont été exportés au format Excel afin de permettre leur intégration dans Power BI. Cette transition entre l'environnement de développement (Python) et l'outil de visualisation (Power BI) garantit une continuité fluide entre les analyses techniques et leur représentation graphique interactive.

Dans Power BI, les fichiers générés ont été structurés selon une architecture en constellation, facilitant une organisation logique, modulaire et cohérente des données. Ce modèle offre une grande flexibilité pour la création de tableaux de bord interactifs et l'exploration multidimensionnelle des indicateurs, comme illustré dans la figure ci-dessous :

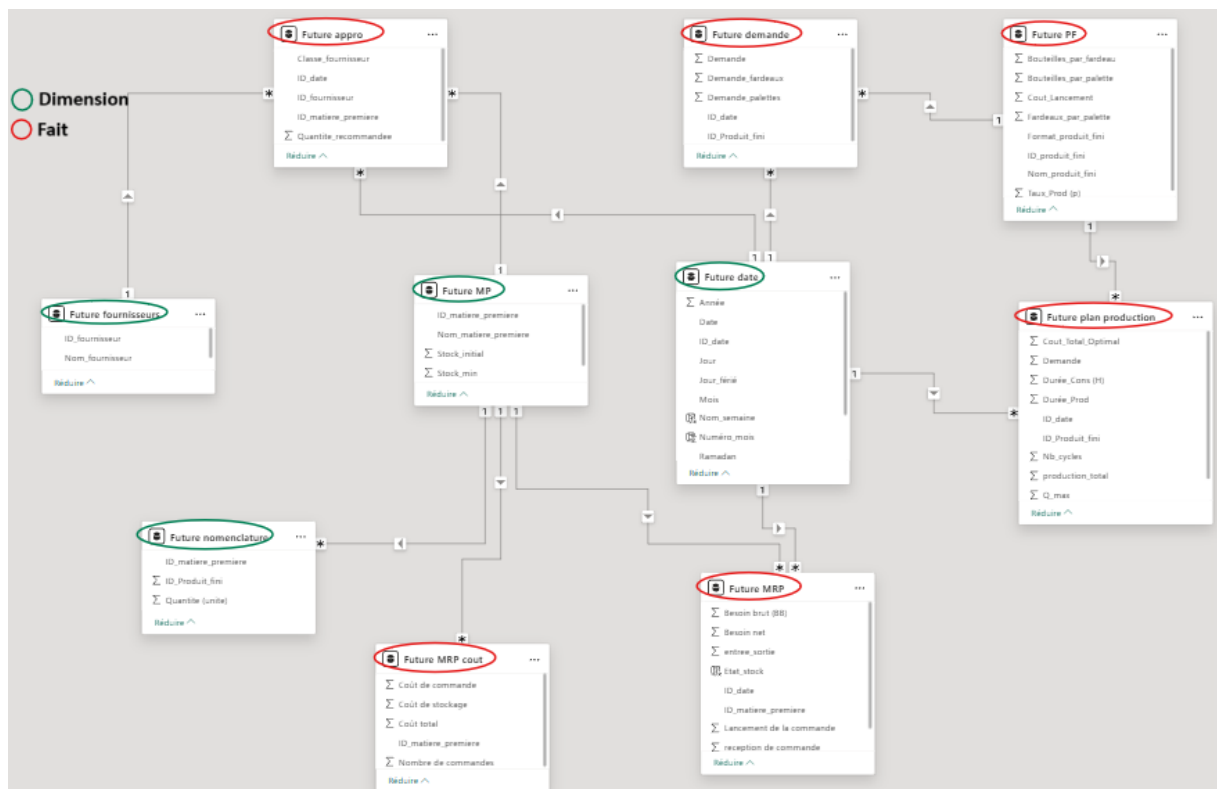


FIGURE 4.38 – Schéma en constellation 2.

Toutefois, nous soulignons que ce pipeline reste, à ce stade, partiellement manuel, notamment en ce qui concerne l'exportation et l'intégration des données. Dans une perspective d'amélioration continue, nous envisageons l'automatisation complète de cette chaîne de traitement à l'aide de solutions ETL. Une telle automatisation permettrait non seulement de garantir une actualisation en temps réel des tableaux de bord, mais également de renforcer la réactivité du processus décisionnel tout en minimisant les risques d'erreurs humaines.

Conclusion

Ce chapitre s'inscrit dans la continuité logique du travail amorcé précédemment. Il applique les fondements théoriques développés au chapitre 2 et met en œuvre la méthodologie détaillée au chapitre 3, à travers une étude de cas menée au sein de l'entreprise Ovitale. Cette étude a porté sur l'optimisation de quatre processus métiers clés, les ventes, la production, l'approvisionnement et la gestion des stocks. L'approche adoptée s'est articulée autour des quatre niveaux d'analyse, permettant d'atteindre les objectifs visés sous des angles complémentaires :

- **L'analyse descriptive** a permis d'établir un état des lieux objectif des performances passées, en s'appuyant sur l'exploration des données historiques et la création de tableaux de bord interactifs.
- **L'analyse diagnostique** a approfondi la compréhension des causes des variations observées, en identifiant les corrélations entre les facteurs internes et externes.
- **L'analyse prédictive** a mobilisé des modèles de ML afin d'anticiper la demande, avec des résultats facilement interprétables grâce à des visualisations dynamiques.
- **L'analyse prescriptive** a transformé ces prédictions en recommandations opérationnelles, en proposant des ajustements concrets concernant les volumes de production, les stratégies d'approvisionnement et les niveaux de stock, à l'aide d'algorithmes d'optimisation.

Cette étude de cas confirme la pertinence d'une approche data-driven pour améliorer l'agilité, la réactivité et la performance des processus métiers. Elle démontre également que l'utilisation combinée de Python, pour l'analyse et la modélisation des données, et de Power BI, pour leur visualisation interactive, permet de convertir les données brutes en outils décisionnels puissants. Cette complémentarité oriente l'entreprise vers un pilotage intelligent, fondé sur les données plutôt que sur l'intuition.

Conclusion générale

Ce travail visait à améliorer les processus métiers clés d'Ovitale (ventes, production, approvisionnement et gestion des stocks) en adoptant une approche data-driven. En s'appuyant sur les apports conjoints de la data science et de la BI, il a démontré comment une PME industrielle algérienne peut valoriser ses données pour accroître son agilité, sa performance opérationnelle et sa compétitivité. La méthodologie déployée s'est structurée autour de quatre axes d'analyse complémentaires : descriptive, diagnostique, prédictive et prescriptive, chacun ayant permis de répondre à des objectifs spécifiques.

Cette étude a permis de poser des fondations solides pour accompagner Ovitale dans sa transition vers une culture pilotée par les données. Elle démontre la pertinence, la faisabilité et la valeur ajoutée d'une stratégie data-driven dans le contexte d'une PME industrielle algérienne, et souligne que l'exploitation intelligente des données constitue aujourd'hui un facteur stratégique essentiel pour renforcer la performance, la résilience et la capacité d'adaptation dans un environnement en perpétuelle évolution.

Cependant, plusieurs limites structurelles et organisationnelles ont été identifiées, ce qui a limité la possibilité d'élargir davantage le périmètre de notre étude. Ces limites concernent notamment :

- Une qualité de données parfois insuffisante, caractérisée par une faible granularité, un volume limité et un accès restreint aux sources d'information.
- L'insuffisance d'infrastructures technologiques adaptées, ne permettant pas une collecte, un traitement et une exploitation fluide, sécurisée et évolutive des données.
- L'absence d'une gouvernance des données formalisée et d'une culture data bien ancrée, limitant la collaboration autour des projets analytiques.

La transformation vers une organisation pleinement data-driven ne se limite pas à la mise en œuvre des quatre axes analytiques. Elle repose sur une approche systémique, intégrant à la fois les volets technologiques, organisationnels, humains et stratégiques.

Pour compléter cette transition et déployer une transformation globale, plusieurs axes d'amélioration doivent être envisagés. Certains relèvent de l'action des équipes opérationnelles, tandis que d'autres nécessitent une vision stratégique à long terme, notamment en matière de politique d'entreprise :

- Centraliser les données dans un environnement unifié et structuré afin de garantir leur accessibilité, cohérence et sécurité.
- Étendre l'approche analytique à d'autres fonctions de l'entreprise telles que le marketing, la logistique ou les ressources humaines.

-
- Intégrer des sources de données enrichies, internes et externes, pour améliorer la profondeur et la précision des analyses.
 - Améliorer les modèles prédictifs afin qu'ils soient plus performants, plus précis et mieux adaptés aux spécificités métiers, en s'appuyant sur des données plus riches, variées et complètes.
 - Automatiser le pipeline analytique (de Python à Power BI) pour accroître la réactivité, la robustesse et la fiabilité des prises de décision.
 - Intégrer l'analyse augmentée, qui combine l'IA et le ML, afin d'automatiser certaines étapes du processus analytique (nettoyage, visualisation. . .), en particulier lorsque l'entreprise commencera à traiter de grandes volumétries de données.
 - Démocratiser l'accès aux données et diffuser une culture data à tous les niveaux de l'organisation, en vulgarisant les concepts clés et en sensibilisant les collaborateurs, même non techniques, à leur rôle dans l'exploitation stratégique de l'information.
 - Constituer des équipes analytiques pluridisciplinaires, regroupant des profils en BI, data science, data engineering et métiers, pour couvrir l'ensemble du cycle de vie des données.
 - Investir dans la structuration des données, le développement des compétences internes, et l'adoption progressive d'outils technologiques avancés en cohérence avec les objectifs stratégiques de l'entreprise.

Bibliographie

- [And15] Carl ANDERSON. *Creating a data-driven organization: Practical advice from the trenches*. " O'Reilly Media, Inc.", 2015.
- [Bog18] Aleksei V BOGOVIZ. « Industry 4.0 as a new vector of growth and development of knowledge economy ». In : *Industry 4.0: Industrial Revolution of the 21st Century*. Springer, 2018, p. 85-91.
- [BM24] Martine BROTKE et Fabian MYRVANG. « Understanding the concept of data-driven organizations: A study on how data-driven approaches impact management accounting ». Mém. de mast. NORWEGIAN SCHOOL OF ECONOMICS, 2024.
- [BN22] P. BUŁA et B. NIEDZIELSKI. *Management, Organisations and Artificial Intelligence: Where Theory Meets Practice*. Routledge Studies in Innovation, Organizations and Technology. Abingdon, Oxon et New York, NY : Routledge, 2022. ISBN : 978-1-032-02582-7.
- [DG20] Mario DÖBLER et Tim GROSSMANN. *The Data Visualization Workshop: An Interactive Approach to Learning Data Visualization*. Packt Publishing Ltd, 2020.
- [For19] Jay E FORTENBERRY. *Optimizing the supply chain*. Business Expert Press, 2019.
- [Gér17] Aurélien GÉRON. *Hands-On machine learning with Scikit-Learn and TensorFlow: concepts, tools*. 2017.
- [Göl17] Philipp GÖLZER. « Big Data in Industrie 4.0-Eine strukturierte Aufarbeitung von Anforderungen, Anwendungsfällen und deren Umsetzung ». Thèse de doct. Dissertation, Erlangen, Friedrich-Alexander-Universität Erlangen-Nürnberg . . . , 2017.
- [Gru19] Joel GRUS. *Data science from scratch: first principles with python*. O'Reilly Media, 2019.
- [NV17] Albert NOGUÉS et Juan VALLADARES. *Business intelligence tools for small companies*. Springer, 2017.
- [Ozd16] Sinan OZDEMIR. *Principles of data science*. Packt Publishing Ltd, 2016.
- [Özü+18] Ayşe Gökçü ÖZÜDÖĞRU et al. « How industry 4.0 changes business: A commercial perspective ». In : *International Journal of Commerce and Finance* 4.1 (2018), p. 84-95.
- [TN20] Eva TYLEČKOVÁ et Darja NOSKIEVIČOVÁ. « The role of big data in Industry 4.0 in mining industry in Serbia ». In : *System Safety: Human-Technical Facility-Environment* 2.1 (2020), p. 166-173.

- [VEP18] Nancy VELÁSQUEZ, Elsa Clara ESTEVEZ et Patricia Mabel PESADO. « Cloud computing, big data and the industry 4.0 reference architectures ». In : *Journal of Computer Science & Technology* 18 (2018).
- [Vor+23] Valentyna VORONKOVA et al. « Digital technology evolution of the industrial revolution from 4g to 5g in the context of the challenges of digital globalization ». In : *TEM Journal* 12.2 (2023), p. 732-742.
- [Zwi22] Tobias ZWINGMANN. *Ai-powered business intelligence*. " O'Reilly Media, Inc.", 2022.

Webographie

- [Act25] ACTIVE TRAIL. *Top 6 des meilleurs logiciels de Marketing Automation en 2025*. Consulté le 01/05/2025. 2025. URL : https://www.activetrail.fr/blog_marketing/marketing_automation_articles_fr/top-5-meilleurs-logiciels-marketing-automation/#4_Quel_est_le_meilleur_logiciel_Marketing_Automation_en_2025.
- [Ana23] Royana ANAND. *Case Studies in Data Analytics: Success Stories*. Consulté le 27/04/2025. 2023. URL : <https://www.geekster.in/articles/case-studies-in-data-analytics-success-stories/>.
- [Ber22] Christina BERNARD. *What Are Some Examples of Top Data-Driven Companies?* Consulté le 27/04/2025. 2022. URL : <https://www.phdata.io/blog/examples-of-data-driven-companies/>.
- [cli24] CLICKUP. *Les 10 meilleurs logiciels d'analyse prédictive pour prendre des décisions basées sur les données*. Consulté le 30/04/2025. 2024. URL : <https://clickup.com/fr-FR/blog/126781/logiciel-d-analyse-predictive>.
- [Dat25] BRIGHT DATA. *Les 10 meilleurs services de collecte de données en 2025*. Consulté le 01/05/2025. 2025. URL : <https://brightdata.fr/blog/donnees-web/best-data-collection-services>.
- [Dat24] DATA SCIENTEST. *Les métriques en Machine Learning : comment évaluer les performances d'un modèle ?* Consulté le 13/06/2025. 2024. URL : <https://datascientest.com/metriques-en-machine-learning>.
- [Dev24] Google DEVELOPERS. *Accuracy, Precision and Recall | Machine Learning Crash Course*. Consulté le 13/06/2025. 2024. URL : <https://developers.google.com/machine-learning/crash-course/classification/accuracy-precision-recall?hl=fr>.
- [Gui25] Julia GUINAMARD. *Top 10 des meilleurs logiciels CRM en 2024*. Consulté le 01/05/2025. 2025. URL : <https://siecledigital.fr/2025/01/05/top-10-des-meilleurs-logiciels-crm-en-2024/>.
- [ins23] INSIGHT SOFTWARE. *Comparing Descriptive, Predictive, Prescriptive, and Diagnostic Analytics*. Consulté le 22/04/2025. 2023. URL : <https://insightsoftware.com/blog/comparing-descriptive-predictive-prescriptive-and-diagnostic-analytics/#what-are-the-4-types-of-data-analysis%3f>.

- [IT23] Inventiv IT. *Comment mettre en place une stratégie data driven ?* Consulté le 05/05/2025. 2023. URL : <https://inventiv-it.fr/comment-mettre-en-place-strategie-data-driven/>.
- [Jou24] Analyst JOURNEY. *Data-Driven Decision-Making Case Studies: Insights from Real-World Examples*. Consulté le 27/04/2025. 2024. URL : <https://www.analyst-journey.com/post/data-driven-decision-making-case-studies-insights-from-real-world-examples>.
- [Ovi] OVITALE. *Ovitale - Eau de source et boissons gazeuses*. Consulté le 05/05/2025. URL : <https://www.ovitale.dz/>.
- [Pro25] PROJEXION. *Devenir une entreprise data driven : Comment ? Quels conseils ?* Consulté le 29/04/2025. 2025. URL : <https://www.projexion.com/enjeux/devenir-une-entreprise-data-driven/>.
- [Sal24] SALESFORCE. *Qu'est-ce qu'une approche Data Driven ? Définition et avantages*. Consulté le 01/05/2025. 2024. URL : <https://www.salesforce.com/fr/marketing/data-driven-marketing/>.
- [Ser22] SERTIS. *5 Interesting Case Studies of Companies Using Data*. Consulté le 27/04/2025. 2022. URL : <https://sertiscorp.medium.com/5-interesting-case-studies-of-companies-using-data-69e1124aa5c7>.

Résumé

Ce mémoire porte sur l'optimisation des processus métiers chez Ovitala à travers une approche data-driven, dans le but de renforcer l'efficacité décisionnelle et opérationnelle de l'entreprise. Face à une croissance constante des données et à des décisions souvent basées sur l'intuition, il propose une démarche structurée alliant Business Intelligence et Data Science.

La méthodologie adoptée repose sur les quatre types d'analyse de données : descriptive, diagnostique, prédictive et prescriptive. L'étude s'appuie sur des données internes d'Ovitala (ventes, production, stock, approvisionnement), enrichies par des sources externes (données météorologiques, jours fériés, événements), collectées via web scraping et interface de programmation d'applications (API). Les outils mobilisés sont Python et Power BI.

L'analyse descriptive a permis de mieux comprendre les dynamiques internes de l'entreprise, tandis que l'analyse diagnostique a aidé à identifier les causes sous-jacentes de certains dysfonctionnements. L'analyse prédictive, fondée sur des modèles de machine learning, a permis d'anticiper l'évolution de la demande. Quant à l'analyse prescriptive, elle a mobilisé des algorithmes d'optimisation en vue d'améliorer la performance de la chaîne logistique. Enfin, des tableaux de bord interactifs ont été développés afin de soutenir la prise de décision en temps réel.

Ce travail montre que la convergence entre la Business Intelligence et la Data Science peut transformer une entreprise traditionnelle en une organisation data-driven, capable d'anticiper les tendances, d'optimiser ses ressources et d'améliorer sa performance globale de manière intelligente et proactive. Il constitue ainsi une première étape vers la mise en place d'un système décisionnel intelligent, répliquable à d'autres départements de l'entreprise.

Mots-clés : data-driven, Ovitala, optimisation, informatique décisionnelle, science des données, apprentissage automatique, prise de décision.

Abstract

This document focuses on optimizing business processes at Ovitalé through a data-driven approach, with the objective of enhancing the company's decision-making and operational efficiency. In response to the growing volume of data and the frequent reliance on intuition-based decisions, it proposes a structured methodology combining Business Intelligence and Data Science.

The approach is based on four key types of data analysis : descriptive, diagnostic, predictive, and prescriptive. The study uses internal data from Ovitalé (sales, production, stock, supply), enriched with external sources (weather data, public holidays, events), collected via web scraping and Application Programming Interfaces (APIs). The main tools used are Python and Power BI.

Descriptive analysis helped better understand internal dynamics, while diagnostic analysis identified the root causes of certain inefficiencies. Predictive analysis, built on machine learning models, enabled demand forecasting. Prescriptive analysis applied optimization algorithms to improve supply chain performance. Finally, interactive dashboards were developed to support real-time decision-making.

This work illustrates how the convergence of Business Intelligence and Data Science can transform a traditional company into a data-driven organization, capable of anticipating trends, optimizing its resources, and improving overall performance in a smart and proactive manner. It represents a foundational step toward the implementation of an intelligent decision-making system, scalable across other departments within the company.

Keywords : data-driven, Ovitalé, optimization, business intelligence, data science, machine learning, decision-making.

