



Mastère Chef de Projet Data et Intelligence Artificielle

Année 2023-2024

Projet d'étude M2
Mémoire de Recherche

Développement d'une Pipeline de Données Automatisée : Extraction, Transformation et Visualisation via Tableau

Réalisé par :

VERA COSIO Ray Leonardo

M2 LM DS1 Paris



Table de matières

Résumé.....	1
Mots clés.....	1
Abstract.....	2
Key words.....	2
1. Introduction.....	3
1.1 Contexte et Pertinence de l'Étude.....	3
1.2 Problématique et Objectifs.....	4
1.3 Justification et Importance du Projet.....	5
1.4 Cadre Conceptuel.....	6
2. Méthodologie et Approche Technique.....	7
2.1 Collecte et Extraction des Données.....	7
2.2 Préparation et Nettoyage des Données.....	12
2.3 Choix des Outils et Justification (Tableau, BigQuery, Excel, Python).....	12
2.4 Conception de la Pipeline Automatisée.....	14
3. Développement et Implémentation de la Solution.....	17
3.1 Construction de la Pipeline : Extraction depuis Oracle Essbase et BigQuery.....	17
3.2 Transformation des Données avec Tableau Prep et Python.....	21
3.3 Chargement Automatisé des Données de BigQuery vers Tableau.....	25
3.4 Création de Dashboards Dynamiques avec Tableau.....	25
3.5 Automatisation et Intégration Continue.....	30
4. Évaluation de la Solution et Résultats.....	32
4.1 Performance et Fiabilité de la Pipeline.....	32
4.2 Comparaison avec l'Approche Traditionnelle (SmartView Excel).....	34
4.3 Bénéfices pour le Processus Décisionnel Financier.....	35
4.4 Retour d'Expérience des Utilisateurs Financiers.....	37
5. Mise En Place Du Projet «Solution ».....	39
5.1 Finalité Du Projet.....	39
5.2 Liste Des Taches A Effectuer.....	40
5.3 Durée Du Projet, Planification Et Jalons.....	40
6. Perspectives d'Amélioration et Extension.....	43
6.1 Optimisation Continue et Scalabilité.....	43
6.2 Intégration de Nouvelles Sources de Données.....	43
6.3 Automatisation Avancée avec Machine Learning.....	44

7. Réalisation d'un Business Plan	45
7.1 Évaluation Financière du Projet.....	46
7.2 Perspectives de Gestion et Analyses Financières.....	47
8. Conclusions et Perspectives Future	50
8.1 Synthèse des Résultats et Conclusions.....	50
8.2 Contribution du Projet à l'Entreprise.....	50
8.3 Perspectives de Développement Futur.....	51
9. Bibliographie	52
10. Annexe	55

RÉSUMÉ

L'optimisation et l'automatisation de la gestion des données financières sont des enjeux majeurs dans le contexte actuel des entreprises, où la rapidité et la précision des informations sont essentielles pour une prise de décision éclairée. Ce projet aborde cette problématique en proposant une solution innovante basée sur la mise en place d'une pipeline de données automatisée, permettant l'extraction, la transformation et la visualisation dynamique des données financières à partir de sources telles qu'Oracle Essbase.

Dans cette étude, nous décrivons la conception et l'implémentation d'une pipeline utilisant Excel avec macros, Python, BigQuery, et Tableau pour surmonter les limitations d'accès direct aux données d'Oracle Essbase. Cette approche vise à réduire la dépendance vis-à-vis des équipes techniques et à offrir une plus grande autonomie aux utilisateurs financiers. Nous avons également automatisé le processus pour garantir des mises à jour quotidiennes et une visualisation en temps réel des informations cruciales pour les décideurs.

Notre méthodologie comprend une phase approfondie de collecte et de préparation des données, suivie par la création de flux de travail automatisés pour transformer et charger les données dans BigQuery. Enfin, la création de dashboards dynamiques avec Tableau permet de visualiser les résultats de manière claire et interactive, facilitant ainsi l'analyse financière.

Les résultats obtenus démontrent une amélioration significative en termes de rapidité et de flexibilité dans l'accès et l'analyse des données financières, comparativement à l'approche traditionnelle utilisant SmartView Excel. Ce projet contribue ainsi à transformer le processus décisionnel financier en le rendant plus agile et indépendant des limitations d'accès aux données.

MOTS CLÉS

Automatisation des données, gestion financière, pipeline de données, Oracle Essbase, Tableau, BigQuery, Excel avec macros, Python, visualisation dynamique, dashboards, extraction de données, transformation des données, analyse financière, mise à jour en temps réel, prise de décision.

ABSTRACT

The optimization and automation of financial data management are critical challenges in today's business environment, where speed and accuracy of information are essential for informed decision-making. This project addresses this issue by proposing an innovative solution through the implementation of an automated data pipeline, enabling the extraction, transformation, and dynamic visualization of financial data from sources such as Oracle Essbase.

In this study, we describe the design and implementation of a pipeline utilizing Excel with macros, Python, BigQuery, and Tableau to overcome the limitations of direct access to Oracle Essbase data. This approach aims to reduce dependency on technical teams and provide greater autonomy to financial users. We have also automated the process to ensure daily updates and real-time visualization of critical information for decision-makers.

Our methodology includes an in-depth phase of data collection and preparation, followed by the creation of automated workflows to transform and load data into BigQuery. Finally, the creation of dynamic dashboards with Tableau allows for clear and interactive visualization of the results, facilitating financial analysis.

The results obtained demonstrate a significant improvement in the speed and flexibility of accessing and analyzing financial data compared to the traditional approach using SmartView Excel. This project thus contributes to transforming the financial decision-making process by making it more agile and independent of data access limitations.

KEYWORDS

Data automation, financial management, data pipeline, Oracle Essbase, Tableau, BigQuery, Excel with macros, Python, dynamic visualization, dashboards, data extraction, data transformation, financial analysis, real-time updates, decision-making.

1. INTRODUCTION:

Dans le domaine de la gestion financière d'entreprise, la rapidité et l'efficacité des processus décisionnels sont devenues des enjeux cruciaux. Les données financières, souvent vastes et complexes, doivent être traitées et analysées de manière fluide pour permettre aux décideurs de réagir rapidement aux fluctuations du marché et aux changements économiques. Cependant, l'accès direct et efficace aux données critiques est souvent limité par des contraintes techniques et organisationnelles, comme c'est le cas dans notre entreprise, Rail Logistics Europe (RLE), une entité de la SNCF.

Au sein du service financier de RLE, l'accès aux données essentielles, telles que celles contenues dans le cube multidimensionnel d'Oracle Essbase appelé "RAILF", se fait principalement via SmartView Excel. Bien que cet outil permette l'extraction et la manipulation des données, il présente des limitations en termes de flexibilité et d'autonomie des utilisateurs. En effet, la nécessité de passer par l'équipe MOA pour la création et la mise à jour des tableaux de données introduit des retards et des inefficacités qui entravent la fluidité du processus décisionnel, particulièrement pour des analyses critiques comme l'évaluation de la marge contributive des différentes entités.

Ce projet vise à surmonter ces limitations en proposant une solution innovante : la mise en place d'un pipeline de données automatisée, reliant Oracle Essbase, Excel avec macros, Python, BigQuery, et Tableau. Ce pipeline permettra un accès plus direct et autonome aux données, tout en automatisant les mises à jour et en permettant une visualisation dynamique des informations financières essentielles. Les graphiques générés offrent déjà aux Directeurs financiers une vue distincte de leurs données, leur permettant de prendre des décisions stratégiques telles que la réaffectation de contrats ou l'évaluation des secteurs en fonction de métriques clés comme l'EBITDA, le ROC, et le chiffre d'affaires (CA).

L'importance de cette étude réside dans la transformation qu'elle promet d'apporter à la gestion des données financières chez RLE. En rendant le processus plus agile et indépendant des contraintes techniques actuelles, ce projet vise à améliorer la qualité et la rapidité des analyses financières, offrant ainsi une base plus solide pour la prise de décisions stratégiques.

1.1 CONTEXTE ET PERTINENCE DE L'ÉTUDE

La gestion des données financières joue un rôle central dans les entreprises modernes, où la rapidité et l'exactitude de l'information sont des impératifs pour soutenir les décisions

stratégiques. Chez Rail Logistics Europe (RLE), l'accès aux données critiques est restreint par la nécessité de passer par SmartView Excel pour extraire des informations du cube multidimensionnel "RAILF" d'Oracle Essbase. Cette méthode, bien qu'efficace pour certaines tâches, se révèle moins adaptée pour répondre rapidement aux besoins d'analyse en temps réel, ce qui complexifie la prise de décisions financières, notamment en ce qui concerne l'analyse de la marge contributive.

Dans ce contexte, ce projet se positionne comme une réponse directe aux défis actuels du service financier de RLE. L'intégration de technologies avancées telles que Python, BigQuery, et Tableau offre la possibilité d'automatiser les flux de données depuis l'extraction jusqu'à la visualisation finale, réduisant ainsi la dépendance vis-à-vis des processus manuels et des interventions techniques. Cette automatisation permet aux Directeurs financiers de disposer d'outils plus performants pour évaluer des secteurs en fonction de l'EBITDA, du ROC, du CA, et pour formuler des recommandations stratégiques, telles que la réallocation des ressources ou l'optimisation des investissements.

La pertinence de ce projet réside donc dans sa capacité à transformer les pratiques actuelles de gestion des données chez RLE, en rendant le processus plus autonome et réactif, tout en fournissant des informations critiques de manière plus rapide et plus efficace. Cela renforce la compétitivité de l'entreprise dans un marché en constante évolution.

1.2 PROBLEMATIQUE ET OBJECTIFS

Problématique :

Dans le contexte actuel de la gestion financière, la complexité croissante des systèmes d'information et la volumétrie des données à traiter posent des défis importants pour les entreprises. Le processus traditionnel d'extraction et d'analyse des données, notamment via des outils comme SmartView Excel connectés à Oracle Essbase, montre des limites en termes de rapidité, de fiabilité et de scalabilité. Ces inefficacités peuvent entraîner des retards dans la prise de décision, des erreurs dans les analyses financières, et une dépendance excessive vis-à-vis des équipes techniques pour l'accès aux données.

Dans un environnement où la prise de décision doit être de plus en plus rapide et basée sur des données fiables, ces limitations deviennent critiques. Il est donc nécessaire d'explorer et de mettre en place des solutions innovantes qui permettent non seulement de surmonter ces défis,

mais aussi de transformer le processus de gestion des données financières en une démarche plus agile et autonome.

Objectifs :

Ce mémoire de recherche se donne pour objectif de développer une solution automatisée de gestion et d'analyse des données financières, intégrant des technologies avancées de Business Intelligence (BI) et d'intelligence artificielle (IA). Les objectifs spécifiques de cette recherche sont :

- **Évaluer l'efficacité des méthodes traditionnelles** d'extraction et d'analyse des données par rapport aux nouvelles solutions technologiques, en mettant l'accent sur l'amélioration du temps de traitement, de la fiabilité et de la précision des données.
- **Développer et tester un pipeline automatisé** de traitement des données basé sur l'intégration d'outils tels que Tableau, BigQuery, Python, et Oracle Essbase, afin de réduire les délais de traitement de 85 à 90 % et d'améliorer la qualité des données disponibles pour les analyses financières.
- **Explorer l'application des technologies IA et BI** dans le contexte de la finance d'entreprise, en mettant l'accent sur la promotion de leur utilisation pour l'analyse des données et la prise de décision stratégique.
- **Proposer des perspectives d'évolution et de scalabilité** pour cette solution, afin qu'elle puisse s'adapter aux besoins futurs de l'entreprise, notamment en termes d'intégration de nouvelles sources de données et d'augmentation des volumes de données à traiter.

1.3 JUSTIFICATION ET IMPORTANCE DU PROJET

Dans le contexte actuel de la gestion financière chez Rail Logistics Europe (RLE), l'accès rapide et autonome aux données critiques est une nécessité pour maintenir une compétitivité élevée. Ce projet vise à répondre à cette exigence en mettant en place une pipeline de données automatisée, capable de surmonter les limitations imposées par l'utilisation de SmartView Excel

pour l'extraction des données d'Oracle Essbase. La justification de ce projet repose sur plusieurs aspects clés.

Premièrement, il permet d'améliorer significativement l'efficacité opérationnelle du service financier, en réduisant la dépendance vis-à-vis de l'équipe MOA pour l'accès aux données et en accélérant le processus d'analyse. Deuxièmement, en automatisant la collecte et la transformation des données, ce projet offre aux décideurs financiers des informations actualisées en temps réel, ce qui est crucial pour prendre des décisions stratégiques éclairées. Enfin, l'intégration de technologies comme Python, BigQuery, et Tableau renforce l'agilité et la flexibilité de l'entreprise, la préparant à répondre rapidement aux dynamiques changeantes du marché. Ce projet est donc non seulement une réponse aux défis techniques actuels, mais aussi un vecteur d'innovation et de compétitivité pour RLE.

1.4 CADRE CONCEPTUEL

Le cadre conceptuel de ce projet est fondé sur l'intégration de technologies avancées et une approche méthodologique rigoureuse pour optimiser la gestion des données financières. Au cœur de cette démarche se trouve l'automatisation des processus de collecte, de transformation et de visualisation des données. Cette approche repose sur une combinaison d'outils, notamment Oracle Essbase pour le stockage des données, Excel avec macros pour l'extraction initiale, Python et BigQuery pour le traitement et la manipulation des données, et enfin Tableau pour la création de dashboards dynamiques. Le cadre conceptuel est guidé par le besoin d'efficacité, de réactivité et d'autonomie dans la gestion des données financières.

En intégrant ces technologies de manière cohérente, le projet vise à créer un pipeline de données automatisée qui non seulement améliore la rapidité et la précision des analyses financières, mais qui permet également de dépasser les limitations actuelles de l'accès aux données. Ce cadre conceptuel offre ainsi une base solide pour repenser et améliorer les pratiques de gestion des données chez RLE, tout en fournissant aux décideurs les outils nécessaires pour naviguer dans un environnement économique complexe et en constante évolution.

2 METHODOLOGIE ET APPROCHE TECHNIQUE

Dans cette section, nous décrivons en détail la méthodologie adoptée pour la mise en place de la pipeline de données automatisée et l'approche technique sous-jacente à ce projet. L'objectif principal est de créer un flux de travail fluide et efficace, permettant l'extraction, la transformation, et la visualisation dynamique des données financières à partir d'Oracle Essbase, en utilisant une combinaison d'outils tels qu'Excel avec macros, Python, BigQuery, et Tableau.

Notre méthodologie s'articule autour de plusieurs étapes clés. Tout d'abord, nous procédons à l'extraction des données depuis Oracle Essbase à l'aide de SmartView Excel, où des macros spécifiques sont développées pour automatiser et standardiser ce processus. Ensuite, les données extraites sont transformées et chargées dans BigQuery via des scripts Python, qui permettent une manipulation avancée des données, notamment leur nettoyage, agrégation, et structuration en vue de leur exploitation ultérieure. Ces scripts sont conçus pour être exécutés automatiquement, garantissant ainsi une mise à jour régulière des données.

Enfin, les données traitées sont visualisées dans Tableau, où des dashboards dynamiques sont créés pour offrir aux décideurs financiers une vue claire et interactive des informations critiques. Cette approche technique, centrée sur l'automatisation et l'intégration des outils, est essentielle pour atteindre l'objectif du projet : rendre la gestion des données plus agile, autonome et réactive, tout en fournissant des analyses financières précises et à jour.

2.1 COLLECTE ET EXTRACTION DES DONNEES

La première étape essentielle de notre méthodologie consiste à collecter et extraire les données financières nécessaires pour ce projet. Les données disponibles couvrent la période de janvier 2023 à 2024 et concernent les entités incluses dans le cube multidimensionnel "RAILF" d'Oracle Essbase. Parmi ces entités, on retrouve des exemples tels que CTI (Captrain Italie), CTE (Captrain Espagne), CTD (Captrain Allemagne), ainsi que d'autres entités supplémentaires.

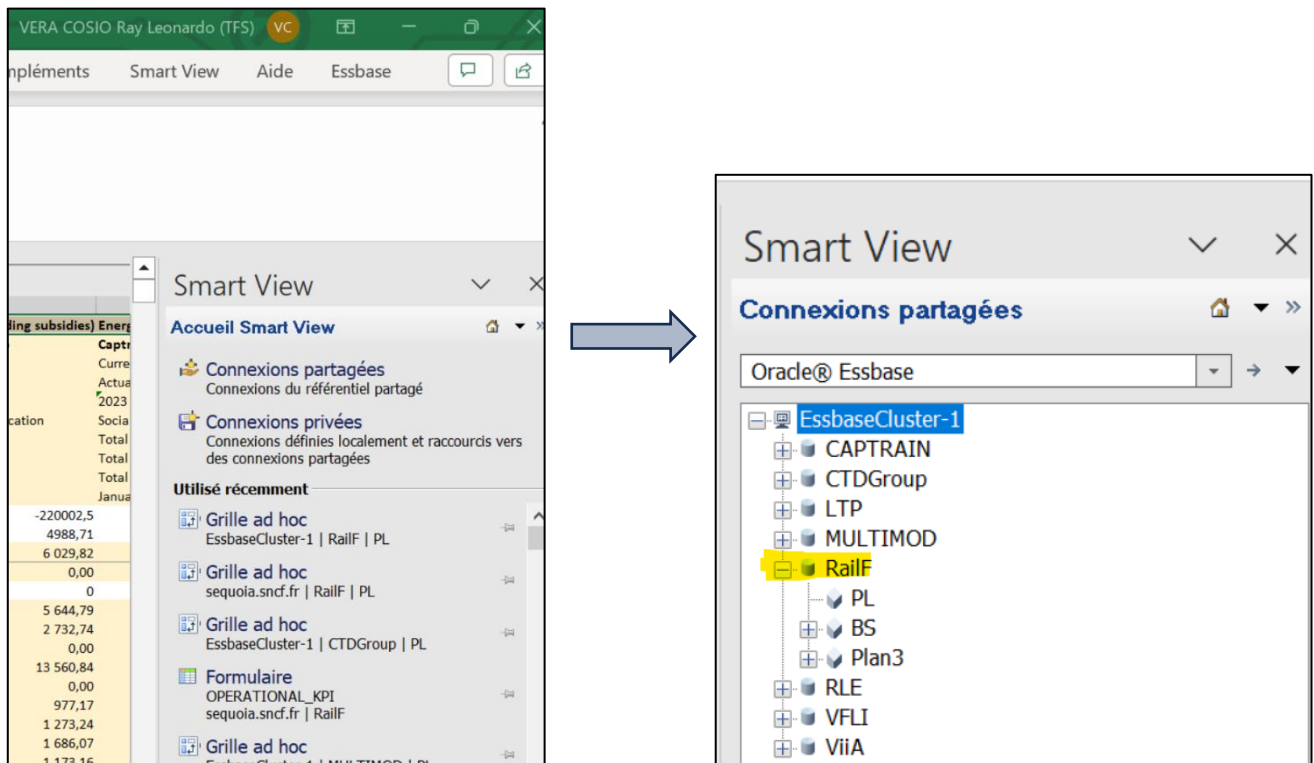


Figure 1 : Panneau Smart-View Excel avec les cubes multidimensionnels

Pour ce projet, les données extraites sont directement liées à la marge contributive, ce qui implique de comprendre en détail la cascade financière associée à cette étude. Le tableau comprend plusieurs colonnes cruciales, telles que : le chiffre d'affaires total (TOTAL Chiffres d'Affaire), le chiffre d'affaires net (TOTAL NET), les coûts opérationnels directs, la marge contributive (Contribution Margin), les coûts opérationnels indirects, et finalement l'EBIT. Une dernière colonne, appelée "Must equal Zero", sert de contrôle. Au total, nous disposons de 19 colonnes, affichant des données pour chaque contrat (WBS ou Work Breakdown Structure) à une date précise (mois, trimestre, semestre, année, etc.).

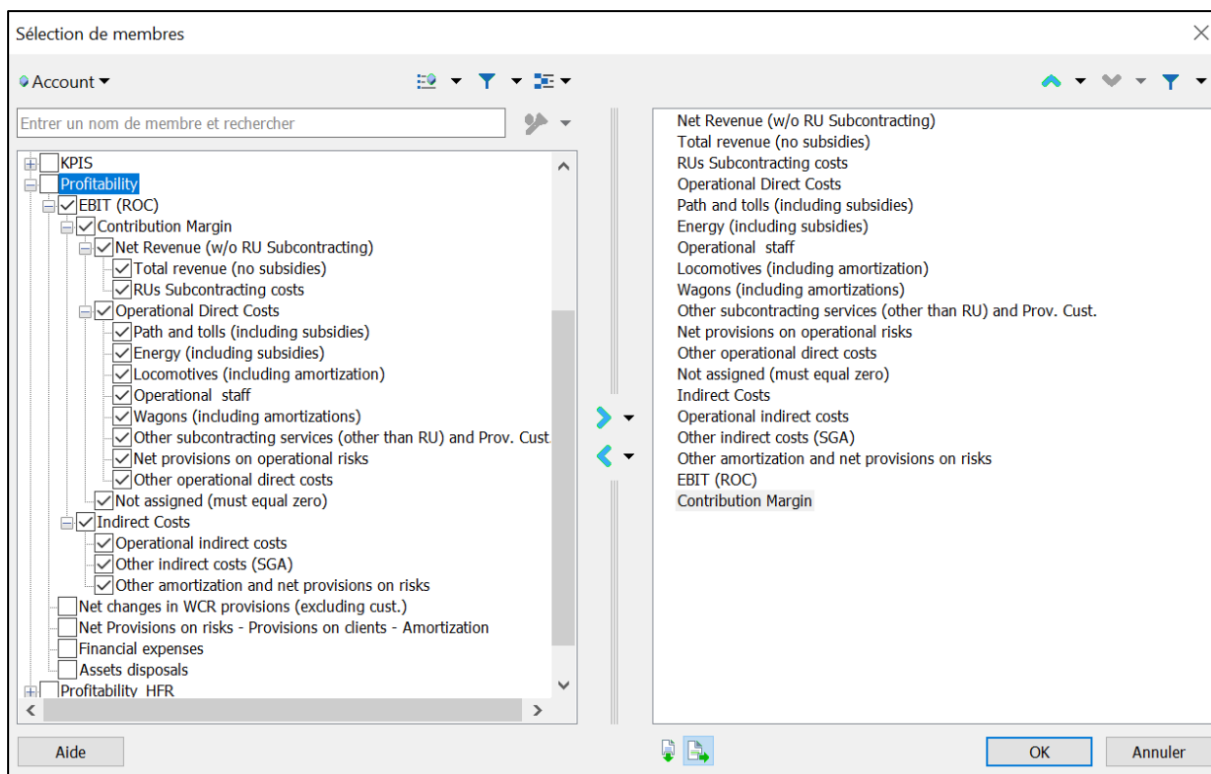


Figure 2 : Sélection de membres à extraire par entité pour le projet « Profitabilité »

Dans ce contexte, nous avons récupéré les données sous forme d'un fichier Excel avec extension « .xlsm », où chaque feuille Excel correspond à une entité différente, regroupant les données de l'année entière avec un détail mensuel. Ainsi, nous disposons d'un aperçu mensuel détaillé de la P&L (Profit and Loss) pour les entités du cube "RAILF". Ce processus d'extraction nécessite deux étapes distinctes : d'abord pour l'année 2023, puis pour l'année 2024.

	A	B	C	D	E	F	G	H	I
1			Total revenue (no subsidies)	RUs Subcontracting costs	Net Revenue (w/o RU Subcontracting)	Path and tolls (including subsidies)	Energy (including subsidies)	Locomotives (including amortization)	Operational staff
2			CAPTRAIN ESPANA,S.A.U.	CAPTRAIN ESPANA,S.A.U.	CAPTRAIN ESPANA,S.A.U.	CAPTRAIN ESPANA,S.A.U.	CAPTRAIN ESPANA,S.A.U.	CAPTRAIN ESPANA,S.A.U.	CAPTRAIN ESPANA,S.A.U.
3			Current version	Current version	Current version	Current version	Current version	Current version	Current version
4			Actual	Actual	Actual	Actual	Actual	Actual	Actual
5			2024	2024	2024	2024	2024	2024	2024
6			Social GAAP after allocation	Social GAAP after allocation	Social GAAP after allocation	Social GAAP after allocation	Social GAAP after allocation	Social GAAP after allocation	Social GAAP after allocation
7			Total functions	Total functions	Total functions	Total functions	Total functions	Total functions	Total functions
8			Total Profit Center	Total Profit Center	Total Profit Center	Total Profit Center	Total Profit Center	Total Profit Center	Total Profit Center
9			Total Partner	Total Partner	Total Partner	Total Partner	Total Partner	Total Partner	Total Partner
10			January	January	January	January	January	January	January
11	Total_Analytic	Analytic Total	5270862,43	-119121,4	5151741,03	-227774,87	-1045017,43	-1485710,55	-1331001,2
12	SP01NAC00	SP01NAC00 - Transport National	0	0	0	0	0	0	0
13	SP01NEWTR1	SP01NEWTR1 - New National_1	0,00	0,00	0,00	0,00	0,00	0,00	0,00
14	SP01NEWTR2	SP01NEWTR2 - New National_2	0,00	0,00	0,00	0,00	0,00	0,00	0,00
15	SP01CAFMAQ	SP01CAFMAQ - CAF - Homologation	0,00	0,00	0,00	0,00	0,00	0,00	0,00
16	SP01CAFPER	SP01CAFPER - CAF - Servicios de	0,00	0,00	0,00	0,00	0,00	0,00	0,00
17	SP01CARGBM	SP01CARGBM - Cargill - Puerto B	0,00	0,00	0,00	0,00	0,00	0,00	0,00
18	SP01CELSAS	SP01CELSAS - Celsa - Silla-Castell	0,00	0,00	0,00	0,00	0,00	0,00	0,00
19	SP01ENDESA	SP01ENDESA - Endesa - Aviles	0,00	0,00	0,00	0,00	0,00	0,00	0,00
20	SP01GMFTRA	SP01GMFTRA - GMF - Traslado d	0,00	0,00	0,00	0,00	0,00	0,00	-0,00
21	SP01STEELT	SP01STEELT - Steeltrack	0,00	0,00	0,00	0,00	0,00	0,00	0,00
22	SP01TRAVIL	SP01TRAVIL - Tramesa - Gijón-Vil	43 329,51	0,00	43 329,51	-957,92	-17 959,67	-14 979,96	-5 841,3
23	SP01PECOVA	SP01PECOVA - Pecovasa - Venta	0,00	0,00	0,00	0,00	0,00	0,00	0,00
24	SP01UTEROV	SP01UTEROV - UTE Rover - Condu	0,00	0,00	0,00	0,00	0,00	0,00	0,00
25	SP01GEFCOB	SP01GEFCOB - Gelfco - Venta de E	0,00	0,00	0,00	0,00	0,00	0,00	0,00
26	SP01TORRAS	SP01TORRAS - Torraspapel - Zarz	0,00	0,00	0,00	0,00	0,00	0,00	0,00
27	SP01VERTBP	SP01VERTBP - VERTEX - Babilafue	0,00	0,00	0,00	0,00	0,00	0,00	0,00
28	SP01DOMMTE	SP01DOMMTE - Homologación de	0,00	0,00	0,00	1 541,33	0,00	0,00	0,00
29	CTE	RTX	CTE	Takargo	Ibercarga	FWD SAS	FWD GMBH	CTD	CTF

Figure 3 : Aperçu partiel des données CTE (CapTrain Espagne) avec quelques colonnes et aussi quelques contrats

Description des données par entité à l'aide de python :

```

✓ 93s Python
<>:4: SyntaxWarning: invalid escape sequence '\p'
<>:4: SyntaxWarning: invalid escape sequence '\p'
C:\Users\User\AppData\Local\Temp\ipykernel_22284\696053736.py:4: SyntaxWarning: invalid escape sequence '\p'
chemin_fichier = 'C:\\LEONARDO\\GEMA IA SCHOOL\\projet 2024\\donneesCubesRaif_MAC.xlsm'
Feuille: CTI, Nombre de contrats: 354, Taille estimée: 1321120 octets
Feuille: RTX, Nombre de contrats: 313, Taille estimée: 599840 octets
Feuille: CTE, Nombre de contrats: 4753, Taille estimée: 8769440 octets
Feuille: Takargo, Nombre de contrats: 498, Taille estimée: 940240 octets
Feuille: Ibercarga, Nombre de contrats: 21, Taille estimée: 62560 octets
Feuille: FWD SAS, Nombre de contrats: 248, Taille estimée: 480240 octets
Feuille: FWD GMBH, Nombre de contrats: 731, Taille estimée: 1368960 octets
Feuille: CTD, Nombre de contrats: 762, Taille estimée: 1426000 octets
Feuille: CTF, Nombre de contrats: 0, Taille estimée: 23920 octets
Taille totale du fichier: 5949518 octets

```

Figure 4: Résultat code python « cm_data.ipynb »

Nous pouvons voir le détail de chaque feuille, qui représente une entité différente. Chaque feuille contient par défaut 19 KPIs du P&L pour chaque mois de l'année, ce qui fait $19 \times 12 = 228$ colonnes au total par feuille. Par exemple, pour CTI (Captrain Italie), nous avons 239 lignes \times 228 colonnes = 54 492 données pour l'année 2023, auxquelles s'ajoutent 6 mois de 2024 ($19 \times 6 \times 239 = 27\,246$), soit **un total de 81 738 données pour l'entité CTI.**

Revenue	Total revenue (no subsidies)
	RUs Subcontracting costs
Net Revenue (w/o RU Subcontracting)	
% Net revenue / Total Revenue	
Operational Direct Costs	Path and tolls (including subsidies)
	Energy (including subsidies)
	Operational staff
	Locomotives (including amortization)
	Wagons (including amortization)
	Other subcontracting services (other than RU) and Prov. Cust.
	Net provisions on operational risks
	Other operational direct costs
Operational Direct Costs	
contribution Margin	
% contribution Margin / Total Revenue	
% contribution Margin / Net Revenue	
Indirect costs	Operational indirect costs
	Other indirect costs (SGA)
	Other amortization and net provisions on risks
Indirect costs	
% Indirect costs / Total Revenue	
EBIT (ROC)	
% EBIT (ROC) / Total Revenue	

Figure 5 : Détails de comment se voit finalement la cascade pivotée pour avoir le vu plus clair du P&L

Description des Colonnes du P&L :

- **Total revenue (no subsidies) (CA - Chiffre d'Affaires) :** Représente le total des revenus générés par l'entreprise avant de prendre en compte les subventions.
- **RUs Subcontracting costs (Cout de sous-traitance) :** Désigne les coûts liés à la sous-traitance, c'est-à-dire le recours à d'autres entreprises pour accomplir certaines tâches.
- **Net Revenue (w/o RU Subcontracting) (% sous-traitance par rapp au CA) :** Représente le chiffre d'affaires net après avoir soustrait les coûts de sous-traitance du chiffre d'affaires total.
- **Path and tolls (including subsidies) (Péages) :** Désigne les coûts liés à l'utilisation des infrastructures routières, y compris les péages, en tenant compte des subventions éventuelles.
- **Energy (including subsidies) (Énergie - diesel et électrique) :** Inclut les coûts énergétiques, y compris les subventions pour le diesel et l'électricité.
- **Locomotives (including amortization) (Coûts de location des locomotives - maintenance) :** Regroupe les coûts de location des locomotives, incluant leur maintenance et amortissement.
- **Operational staff (Charges de personnel - masse salariale) :** Concerne les dépenses liées aux salaires et charges du personnel.
- **Wagons (including amortizations) (Location des wagons - amortissement) :** Désigne les coûts de location des wagons, y compris leur amortissement.
- **Other subcontracting services (other than RU) and Prov. Cust. (Charges sous-traitance - autres) :** Regroupe les coûts d'autres services de sous-traitance, en excluant les coûts de sous-traitance ferroviaire.
- **Net provisions on operational risks (Provisions nettes sur BFR) :** Représente les provisions pour anticiper les risques opérationnels, tels que les litiges.
- **Other operational direct costs (Autres coûts opérationnels directs) :** Inclut les autres coûts directs liés à l'exploitation.
- **Operational Direct Costs (Coût total opérationnel direct) :** La somme totale des coûts directs liés à l'exploitation.
- **Contribution Margin (Marge contributive) :** La différence entre le chiffre d'affaires net et les coûts opérationnels directs.

- **Operational indirect costs (Coûts indirects opérationnels) :** Désigne les coûts opérationnels indirects qui ne peuvent être affectés spécifiquement à des contrats particuliers.
- **Other indirect costs (SGA) (Frais généraux et administratifs) :** Inclut les autres coûts indirects, tels que les frais généraux et administratifs.
- **Other amortization and net provisions on risks (Autres amortissements et provisions sur risques) :** Concerne les autres amortissements et provisions pour risques qui ne peuvent être affectés spécifiquement à des contrats.
- **Indirect Costs (Coûts indirects totaux) :** La somme totale des coûts indirects.
- **EBIT (ROC) (Résultat opérationnel courant) :** Les bénéfices avant intérêts et impôts.

2.2 PRETRAITEMENT ET NETTOYAGE DES DONNEES

Dans le cadre de ce projet, les données extraites via SmartView Excel sont issues directement d'Oracle Essbase, où elles ont été préalablement structurées et validées par l'équipe MOA. Ces données étant déjà organisées et prêtes à l'emploi, le besoin de nettoyage supplémentaire est limité. Cela permet de réduire le temps et les ressources consacrés à cette étape, garantissant ainsi une utilisation plus rapide des informations pour l'analyse financière.

Cependant, une étape de calcul est intégrée dans Tableau Prep Builder pour la création de certains KPIs spécifiques. Ces calculs sont essentiels pour affiner l'analyse des données et fournir des indicateurs clés, qui seront ensuite visualisés dans les dashboards de Tableau Desktop. Le prétraitement des données via Tableau Prep est donc crucial pour garantir que les KPIs soient calculés avec précision et présentés de manière cohérente, permettant ainsi une analyse financière pertinente et fiable.

2.3 CHOIX DES OUTILS ET JUSTIFICATION (TABLEAU, BIGQUERY, EXCEL, PYTHON)

Le choix des outils utilisés dans ce projet a été guidé par plusieurs facteurs, dont les directives de l'équipe, les besoins spécifiques du projet, et l'opportunité d'optimiser la gestion et l'analyse des données financières.

Tableau a été imposé comme solution par notre équipe en raison de la décision stratégique de migrer vers une intégration complète avec Salesforce et ses serveurs. Cette décision a ouvert l'accès à des outils puissants comme Tableau Prep Builder et Tableau Desktop. Tableau Prep Builder, bien que facultatif, s'est révélé essentiel dans notre pipeline pour restructurer et préparer les données de manière efficace avant leur exploitation dans Tableau Desktop. Ce choix a été motivé par la capacité de Tableau Prep à inspecter les données de manière globale, à les nettoyer et à créer des tables précises, facilitant ainsi leur utilisation ultérieure dans Tableau Desktop. Cela a permis de concentrer Tableau Desktop sur la création d'insights dynamiques, tout en exploitant pleinement sa puissance en matière de Business Intelligence (BI) et sa capacité à produire des tableaux de bord accessibles, sécurisés, mis à jour en temps réel, et validés.

BigQuery a été proposé et implémenté de ma part comme solution de Data Warehouse pour héberger les données. Cette décision était stratégique pour assurer un accès fiable aux données en cas de perte de connexion à Oracle Essbase, un problème récurrent lors des périodes critiques de l'année. BigQuery offre ainsi un accès privilégié à l'équipe financière, leur permettant de continuer à travailler avec les données et à exploiter les dashboards explicatifs même en cas de défaillance du système principal.

Excel a été utilisé par l'équipe de contrôle de gestion pour diverses tâches, notamment la création de graphiques et le contrôle des données. Dans le cadre de ce projet, Excel a été intégré au pipeline pour la collecte des données depuis Oracle Essbase via SmartView et l'automatisation de certaines tâches à l'aide de VBA. Excel a également joué un rôle crucial dans la vérification de la qualité des données, un aspect particulièrement sensible en raison de la disponibilité limitée des mises à jour fournies par l'équipe MOA.

Python sur VS Code a été utilisé pour des inspections et analyses spécifiques des fichiers Excel extraits d'Oracle Essbase, notamment pour vérifier la taille des fichiers, le nombre de feuilles, les espaces vides, et les formules présentes. Ces informations ont été utilisées pour alléger les fichiers ou ajuster les données. De plus, Python a été employé via l'API TabPy pour effectuer des restructurations avancées dans Tableau Prep, créant des champs calculés complexes et envisageant des modèles prédictifs selon les besoins des utilisateurs.

2.4 CONCEPTION DE LA PIPELINE AUTOMATISEE

Avant de plonger dans le développement de la solution, il a été nécessaire de planifier soigneusement le pipeline automatisé en tenant compte des outils disponibles et des exigences spécifiques du projet. Cette étape de planification a permis de définir clairement les différentes phases du projet et d'établir un calendrier de travail réaliste.

Processus Général

Le pipeline de données automatisée a été conçue pour garantir une collecte, une transformation, et une visualisation efficace des données financières critiques. Le processus général se déroule comme suit :

- ✓ **Extraction des Données** : Les données sont extraites d'Oracle Essbase via SmartView Excel. Cette extraction se fait de manière semi-automatique à l'aide de macros VBA, garantissant ainsi la qualité des données avant leur transfert vers BigQuery.
- ✓ **Stockage et Préparation des Données** : Une fois extraites, les données sont transférées dans BigQuery, où elles sont stockées de manière sécurisée. BigQuery sert de Data Warehouse, centralisant les données pour assurer un accès rapide et fiable, même en cas de défaillance d'Oracle Essbase.
- ✓ **Transformation des Données** : Les données stockées dans BigQuery sont ensuite transformées et restructurées à l'aide de Tableau Prep Builder. Cette étape permet de préparer les données pour une exploitation optimale dans Tableau Desktop, en créant des tables bien structurées et en calculant des KPIs spécifiques.
- ✓ **Visualisation des Données** : Les données préparées sont importées dans Tableau Desktop, où elles sont utilisées pour créer des dashboards dynamiques. Ces dashboards fournissent des insights clairs et exploitables pour les décideurs financiers, leur permettant de prendre des décisions informées en temps réel.
- ✓ **Automatisation et Mises à Jour** : L'ensemble de le pipeline est automatisé pour garantir des mises à jour régulières des données et des dashboards. Cela permet de maintenir une continuité dans l'analyse des données sans intervention manuelle, assurant ainsi une fiabilité et une efficacité maximales.

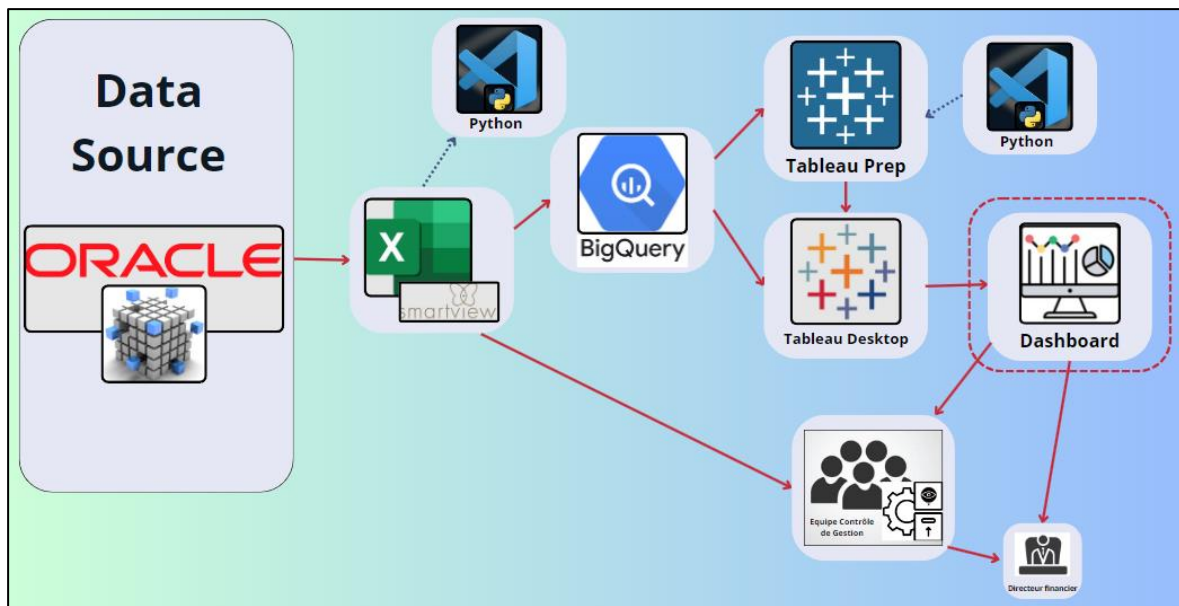


Figure 6 : Pipeline Complète

Élaboration d'une Enquête

L'élaboration de l'enquête a été structurée en trois étapes principales :

- ✓ **Informations recherchées :** L'objectif principal de cette enquête était de recueillir des données financières précises liées aux performances des entités du cube "RAILF" dans Oracle Essbase. Plus précisément, les informations recherchées incluaient des indicateurs financiers tels que la marge contributive, l'EBIT, et d'autres KPIs pertinents pour chaque entité sur les années 2023 et 2024. Ces données devaient être représentatives de la performance économique de chaque entité pour une analyse comparative et prédictive.
- ✓ **Ciblage/échantillon :** Le ciblage de l'enquête s'est concentré sur les entités financières incluses dans le cube multidimensionnel "RAILF", à savoir des entités telles que CTI (Captrain Italie), CTE (Captrain Espagne), et d'autres. L'échantillon comprenait toutes les entités disponibles dans le cube pour garantir une couverture complète et représentative des performances globales. Cette sélection a permis de cibler les contrats WBS les plus significatifs pour obtenir un ensemble de données pertinent.
- ✓ **Mode de recherche :** Le mode de recherche a combiné l'extraction de données via SmartView Excel, l'intégration de ces données dans BigQuery pour un stockage centralisé, et l'utilisation de Tableau pour l'analyse et la visualisation. Cette approche a

permis de structurer l'enquête de manière à ce que les données collectées puissent être transformées en insights exploitables via une pipeline automatisée, garantissant ainsi la continuité et la précision des analyses réalisées.

3. DEVELOPPEMENT ET IMPLEMENTATION DE LA SOLUTION

La réalisation et l'implémentation de ce pipeline peuvent sembler être une étape relativement facile sur un ordinateur personnel. Cependant, parmi les jalons que nous avons définis, nous avons ajouté 3 à 4 jours ouvrables supplémentaires pour chaque partie, en tenant compte non seulement de la difficulté technique, mais aussi du fait que Tableau, Python et BigQuery étaient des outils nouveaux pour ma direction. Cela a nécessité une étape supplémentaire de vérification de la cybersécurité lors de la demande d'installation auprès du service informatique, qui était plus habitué à installer des logiciels bureautiques.

J'ai donc passé un temps considérable à obtenir les logiciels, à faire des demandes de droits d'utilisation pour Tableau, Python et BigQuery. À chaque étape, nous avons rencontré des blocages imprévus, ce qui est normal en entreprise lorsque les outils ne sont pas encore entièrement déployés. Malgré ces défis, j'ai veillé à suivre les bonnes pratiques pour répondre aux besoins des utilisateurs en ce qui concerne le résultat final

3.1 CONSTRUCTION DE LA PIPELINE : EXTRACTION DEPUIS ORACLE ESSBASE ET BIGQUERY

Oracle Essbase

Oracle Essbase a servi comme source principale de données dans ce projet, constituant le point de départ de notre pipeline. Bien que nous n'ayons pas pu intervenir directement sur le programme en raison des restrictions imposées par la direction financière, nous avons pu proposer différentes solutions pour optimiser le processus d'extraction des données.

Initialement, nous avions envisagé de connecter directement le logiciel Tableau à la base de données Oracle Essbase. Cependant, cette idée n'a pas pu être réalisée en raison des limitations d'accès. En réponse à cette contrainte, j'ai proposé d'optimiser les extractions pour ce projet en particulier en utilisant des macros VBA.

Ces macros VBA ont été conçues pour contrôler le fichier d'extraction et recueillir l'ensemble des données nécessaires, par exemple, celles relatives à la Marge Contributive. Les macros permettent de gérer les données par année (en collectant ce qui est nécessaire pour chaque année), de contrôler les données extraites, et d'identifier les incohérences. Une macro spécifique vérifie les données en affichant un indicateur à la fin de chaque colonne : vert si les données

sont correctes, rouge si elles ne le sont pas. Ensuite, une autre macro permet de réinitialiser ces contrôles pour relancer l'extraction des données.

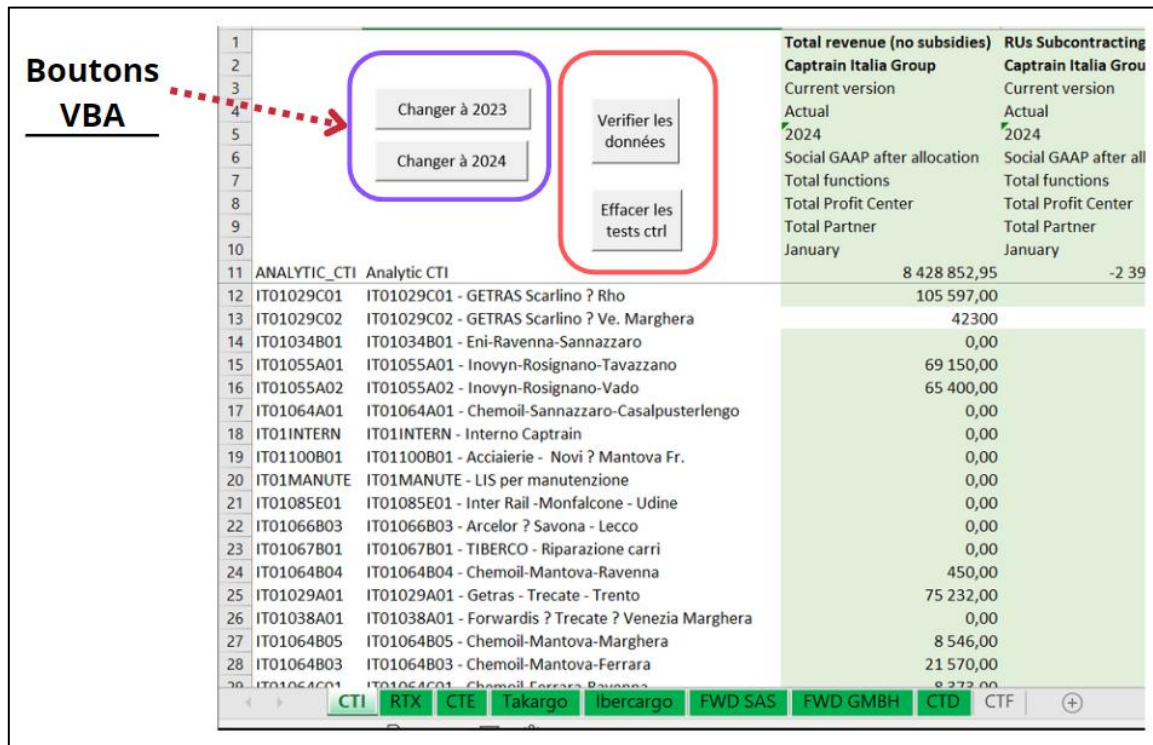


Figure 7 : Capture d'écran boutons VBA (code dans l'annexe)

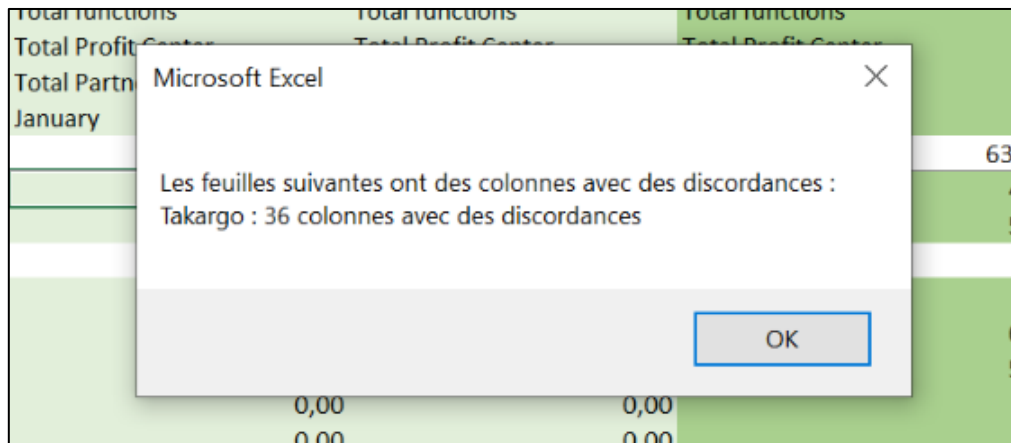


Figure 8 : Résultat de macro VBA qui cherche discordances dans les colonnes des données
On voit que l'entité « Takargo » a 36 colonnes à vérifier.

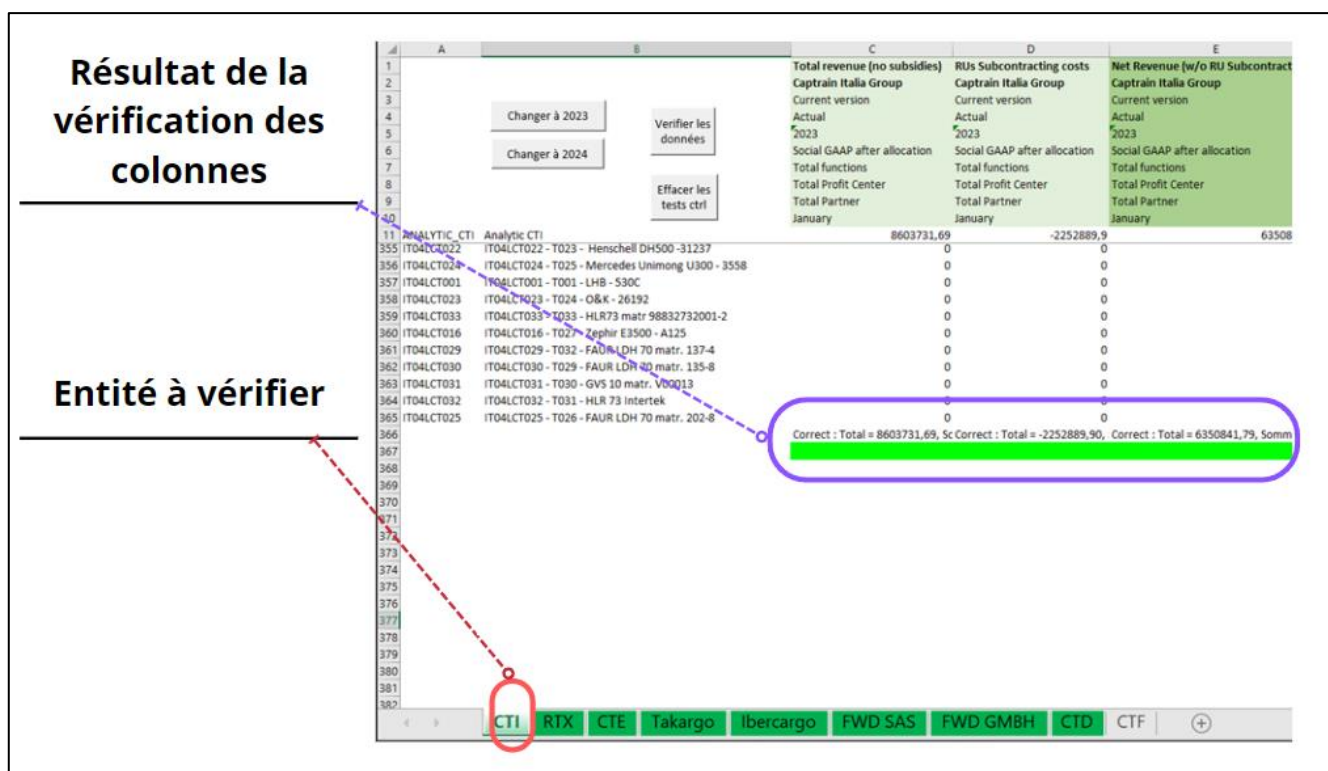


Figure 9 : Résultat de la vérification de CTI, on voit que les données ne sont pas discordantes



Figure 10 : Résultat de la vérification de « Takargo » (Captrain Portugal), on voit que les données son discordantes.

La macro de vérification a été demandée par les utilisateurs, en particulier par l'équipe de contrôle de gestion, pour effectuer un contrôle essentiel des données. Cette étape nous permet

de prendre deux actions : d'abord, notifier l'équipe MOA des incohérences pour qu'elle mette à jour les données, et ensuite, continuer le processus même si certaines données ne sont pas fiables, en fournissant un rapport explicatif à l'équipe financière et au directeur financier sur les éventuelles discordances.

Une fois ces étapes terminées, les données sont prêtes à être transférées vers l'étape suivante.

Big Query

Bien que cette étape soit relativement courte, elle revêt une importance cruciale pour centraliser les tables extraites et les rendre disponibles pour Tableau Prep, Tableau Desktop, ou encore offrir une voie alternative à l'équipe de contrôle de gestion en cas de besoin. BigQuery a également joué un rôle important pour pallier les éventuelles défaillances au niveau de l'équipe MOA en charge d'Oracle Essbase, ainsi qu'assurer la fiabilité des données.

Par exemple, il arrivait parfois que certaines données soient manquantes pour une entité ou que des détails de contrats soient déplacés ou supprimés. Grâce à BigQuery, nous avons pu établir un entrepôt de données sécurisé qui permet un accès continu, au moins hebdomadaire, aux données. De plus, en offrant une vue dynamique avec Tableau, BigQuery s'est révélé être une solution particulièrement intéressante, que nous avons testée avec succès et que l'équipe a fortement appréciée pour son utilité.

Entité à vérifier

Résultat de la requête

Informations sur le job	Résultats	Graphique	JSON	Détails de l'exécution	Graphique d'exécution
Ligne	string_field_0	string_field_1	string_field_2	string_field_3	string_field_4
1	null	null	CAPTRAIN ESPANAS A.U.	CAPTRAIN ESPANAS A.U.	CAPTRAIN ES
2	null	null	Current version	Current version	Current versio
3	null	null	Actual	Actual	Actual
4	null	null	2023	2023	2023
5	null	null	Social GAAP after allocation	Social GAAP after allocation	Social GAAP a
6	null	null	Total functions	Total functions	Total function
7	null	null	Total Profit Center	Total Profit Center	Total Profit Ce
8	null	null	Total Partner	Total Partner	Total Partner
9	null	null	January	January	January
10	null	null	Correct : Total = 0,00, Somme	Correct : Total = 0,00, Somme	Correct : Total
11	Unnamed: 0	Unnamed: 1	Total revenue (no subsidies)	RUS Subcontracting costs	Net Revenue (
12	MET_4	4 Opérations de Proximité	0	0	0
13	New_1	New_1 - BLS - Hupac Luino Nov...	0	0	0
14	New_2	New_2 - Abs-Udine - Udine - Tor...	0	0	0
15	New_3	New_3 - Abs-Siak - Udine/Os...	0	0	0
16	New_4	New_4 - Mis - VE Marghera - Lo...	0	0	0
17	New_5	New_5 - Hitachi - Ptoia - Mad...	0	0	0
18	New_6	New_6 - CD Cargo - Tanisio - O...	0	0	0
19	New_7	New_7 - Sangritana - Fossace...	0	0	0
20	New_8	New_8 - Raitrax-Domo-Sacco...	0	0	0

Figure 11 : Exemple d'entrepôt des données de CTE hébergés sur BigQuery

3.2 TRANSFORMATION DES DONNEES AVEC TABLEAU PREP ET PYTHON

Dans cette section, nous allons détailler les étapes clés de la transformation des données que nous avons réalisées à l'aide de Tableau Prep, en intégrant également des scripts Python via l'API TabPy pour certaines opérations avancées.

A. Récupération des Données depuis BigQuery vers Tableau Prep

Tout d'abord, nous avons extrait les données de chaque feuille CSV importée dans BigQuery, qui correspondait à une entité distincte. Une fois les données chargées dans Tableau Prep, nous avons commencé à organiser les flux pour chaque entité afin d'obtenir une vue de contrôle globale. Cela a impliqué la création de flux de données pour chaque feuille CSV, où nous avons effectué un filtrage initial pour éliminer les données non pertinentes. Nous avons conservé uniquement les 19 indicateurs clés de performance (KPI) du P&L (Profit & Loss) pour chaque mois de l'année 2023, puis pour les six premiers mois de 2024.

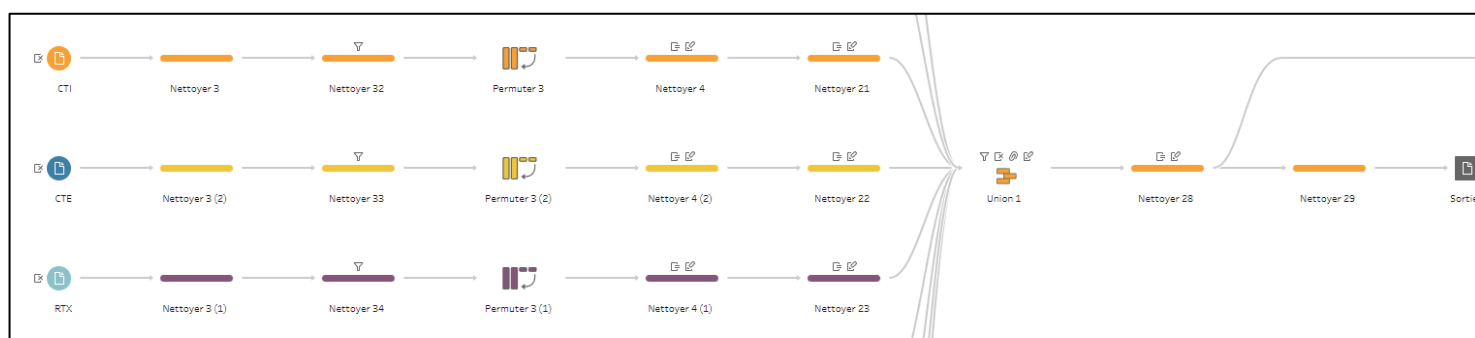


Figure 12 : Extrait de flux Tableau Prep des entités : CTE, CTI et RTX

Abc	Abc	Q	#	#	#	#
F1	F2		Total revenue (no su...	RUs Subcontracting ...	Net Revenue (w/o RU Subcont...	Path a
IT01029C01	IT01029C01 - GETRAS Scarlino ? Rho		46,932	0	46,932	4,988.7
IT01029C02	IT01029C02 - GETRAS Scarlino ? Ve. Marghera		58,665	0	58,665	6,029.8
IT01034B01	IT01034B01 - Eni-Ravenna-Sannazzaro		0	0	0	0
IT01035B03	IT01035B03 - Versalis-Brindisi-Castelguelfo (Spot)		0	0	0	0
IT01055A01	IT01055A01 - Inovyn-Rosignano-Tavazzano		67,150	0	67,150	5,644.7
IT01055A02	IT01055A02 - Inovyn-Rosignano-Vado		50,800	0	50,800	2,732.7
IT01064A01	IT01064A01 - Chemoil-Sannazzaro-Casalpuisterlengo		0	0	0	0
IT01010D01	IT01010D01 - Vtg-Brindisi-Ferrara		0	0	0	0
IT01014A01	IT01014A01 - ALSTOM Vado ? Pistoia		21,430	0	21,430	977.17
IT01085A03	IT01085A03 - INTER RAIL ABS Udine ? Lecco		10,500	0	10,500	1,273.2
IT01085B01	IT01085B01 - INTER RAIL ABS San Stino - Udine		13,950	0	13,950	1,686.0
IT01085C01	IT01085C01 - INTER RAIL ABS Venezia Marghera - Udine		14,400	0	14,400	1,173.1

Figure 13 : Extrait du résultat de la première étape de filtrage dans le générateur Tableau Prep

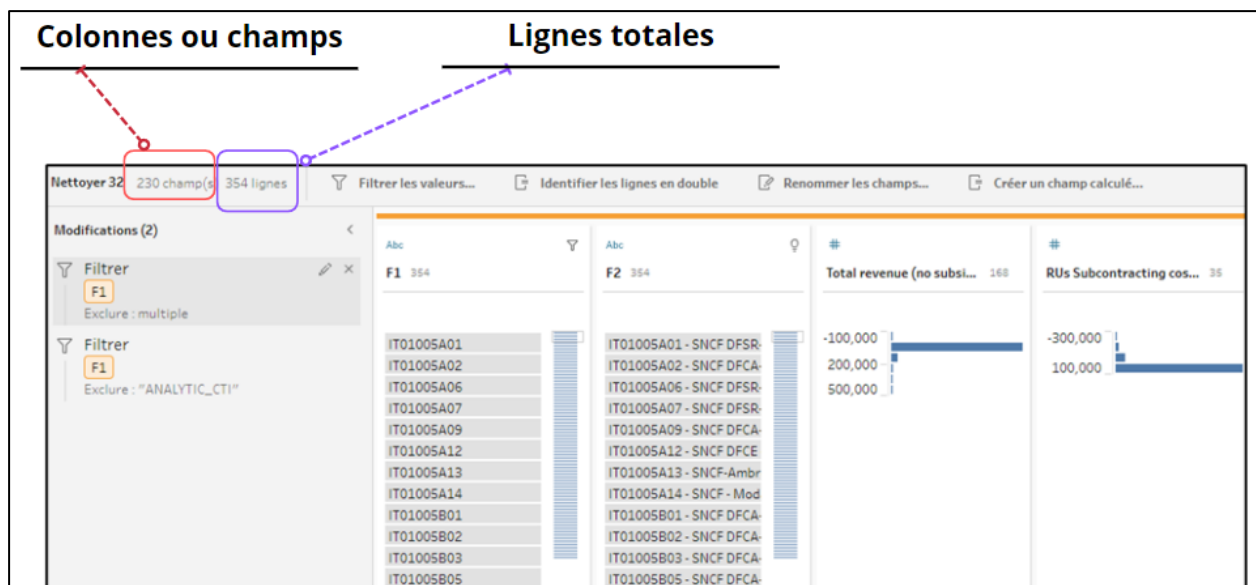


Figure 14 : En affichant les valeurs du profil dans Tableau Prep Builder, nous voyons les 228 colonnes à gauche plus les 2 premières colonnes qui font référence au code et au code du contrat. Voici l'information de CTI, qui correspond aussi à la vérification python que j'avais fait précédemment.

B. Restructuration Initiale des Données

Dans cette première étape de restructuration, nous avons organisé les colonnes de la manière suivante : une colonne pour le code du contrat, une colonne pour le code et le nom du contrat, suivies des 19 colonnes du P&L pour le mois de janvier, puis pour les mois de février à décembre. (19 x 12 colonnes = 228 colonnes)

Ensuite, nous avons pivoté les colonnes du P&L afin de regrouper toutes les valeurs numériques dans une seule colonne, facilitant ainsi les analyses futures.

Résultats de la permutation				
Noms Pivot1 228		#	F1 354	
Contribution Margin		-3,000,000	IT01005A01	IT01005A01 - SNCF DFSR
Contribution Margin 1		2,000,000	IT01005A02	IT01005A02 - SNCF DFCA
Contribution Margin 10		7,000,000	IT01005A06	IT01005A06 - SNCF DFSR
Contribution Margin 11			IT01005A07	IT01005A07 - SNCF DFSR
Contribution Margin 2			IT01005A09	IT01005A09 - SNCF DFCA
Contribution Margin 3			IT01005A12	IT01005A12 - SNCF DFCE
Contribution Margin 4			IT01005A13	IT01005A13 - SNCF-Ambr
Contribution Margin 5			IT01005A14	IT01005A14 - SNCF - Mod
Contribution Margin 6			IT01005B01	IT01005B01 - SNCF DFCA
Contribution Margin 7			IT01005B02	IT01005B02 - SNCF DFCA
Contribution Margin 8			IT01005B03	IT01005B03 - SNCF DFCA
Contribution Margin 9			IT01005B05	IT01005B05 - SNCF DFCA
Noms Pivot1	Permuter1 les valeurs	F1	F2	
Other indirect costs (SGA) 11	0	IT01029C01	IT01029C01 - GETRAS Scarlino ? Rho	
Total revenue (no subsidies)	46,932	IT01029C01	IT01029C01 - GETRAS Scarlino ? Rho	
RUs Subcontracting costs	0	IT01029C01	IT01029C01 - GETRAS Scarlino ? Rho	
Net Revenue (w/o RU Subcontracting)	46,932	IT01029C01	IT01029C01 - GETRAS Scarlino ? Rho	
Path and tolls (including subsidies)	4,988.71	IT01029C01	IT01029C01 - GETRAS Scarlino ? Rho	
Energy (including subsidies)	-108.78	IT01029C01	IT01029C01 - GETRAS Scarlino ? Rho	

Figure 15 : Données et résultat du pivotement des colonnes P&L

C. Transformation des Noms et Création de la Colonne Date

Pour uniformiser les noms après le pivotement, nous avons appliqué une logique conditionnelle qui a permis de nettoyer et numéroter les valeurs de manière cohérente. Cela a été suivi par la création d'une colonne "Date" alignée sur les données P&L, basée sur une correspondance entre les noms des colonnes et les mois de l'année 2023.

Date	Catégorie	Noms Pivot1	Valeurs	Codes	Contrats
01/01/2023	Contribution Margin	Contribution Margin	-3,000,000	IT01005A01	IT01005A01 - SNCF DFRS
01/02/2023	EBIT (ROC)	Contribution Margin 1		IT01005A02	IT01005A02 - SNCF DFCA
01/03/2023	Energy (including subsidi	Contribution Margin 10	2,000,000	IT01005A06	IT01005A06 - SNCF DFRS
01/04/2023	Indirect Costs	Contribution Margin 11		IT01005A07	IT01005A07 - SNCF DFRS
01/05/2023	Locomotives (including ar	Contribution Margin 2	7,000,000	IT01005A09	IT01005A09 - SNCF DFCA
01/06/2023	Net provisions on operati	Contribution Margin 3		IT01005A12	IT01005A12 - SNCF DFCE
01/07/2023	Net Revenue (w/o RU Sub	Contribution Margin 4		IT01005A13	IT01005A13 - SNCF-Ambr
01/08/2023	Not assigned (must equal	Contribution Margin 5		IT01005A14	IT01005A14 - SNCF - Mod
01/09/2023	Operational staff	Contribution Margin 6		IT01005B01	IT01005B01 - SNCF DFCA
01/10/2023	Operational Direct Costs	Contribution Margin 7		IT01005B02	IT01005B02 - SNCF DFCA
01/11/2023	Operational indirect cost	Contribution Margin 8		IT01005B03	IT01005B03 - SNCF DFCA
01/12/2023	Other amortization and n	Contribution Margin 9		IT01005B05	IT01005B05 - SNCF DFCA

Figure 16 : Transformation des noms et l'ajout de la colonne Date (code dans les annexes)

À ce stade, nos données étaient restructurées en six colonnes : Code, Nom du Contrat, P&L, Date, Valeur Numérique, et Catégorie.

D. Création d'une Catégorie de Classification par Corridor

Nous avons ensuite créé une nouvelle catégorie de classification par corridor pour chaque entité. Cette information, obtenue depuis Oracle, n'était pas directement appliquée dans la table Oracle,

Catégorie Contrats	Date	Catégorie	Noms Pivot1	Valeurs	Codes	KPIS
Domestic Corridor	01/01/2023	Contribution Margin	Contribution Margin	-3,000,000	IT01005A01	IT01005A01 - SNCF DFRS
Eastern Corridor	01/02/2023	EBIT (ROC)	Contribution Margin 1		IT01005A02	IT01005A02 - SNCF DFCA
Fleet Management	01/03/2023	Energy (including subsidi	Contribution Margin 10	2,000,000	IT01005A06	IT01005A06 - SNCF DFRS
French Corridor	01/04/2023	Indirect Costs	Contribution Margin 11		IT01005A07	IT01005A07 - SNCF DFRS
North - South Corridor	01/05/2023	Locomotives (including ar	Contribution Margin 2	7,000,000	IT01005A09	IT01005A09 - SNCF DFCA
OTHERS	01/06/2023	Net provisions on operati	Contribution Margin 3		IT01005A12	IT01005A12 - SNCF DFCE
Other Services	01/07/2023	Net Revenue (w/o RU Sub	Contribution Margin 4		IT01005A13	IT01005A13 - SNCF-Ambr
Shunting Contracts	01/08/2023	Not assigned (must equal	Contribution Margin 5		IT01005A14	IT01005A14 - SNCF - Mod
Terminal Services	01/09/2023	Operational staff	Contribution Margin 6		IT01005B01	IT01005B01 - SNCF DFCA
	01/10/2023	Operational Direct Costs	Contribution Margin 7		IT01005B02	IT01005B02 - SNCF DFCA
	01/11/2023	Operational indirect cost	Contribution Margin 8		IT01005B03	IT01005B03 - SNCF DFCA
	01/12/2023	Other amortization and n	Contribution Margin 9		IT01005B05	IT01005B05 - SNCF DFCA

Figure 17 : Données montrant la classification par corridor : « Catégorie Contrats »

car nous n'avions qu'une vue par contrats détaillés. Cette étape a permis de classer les données selon des corridors spécifiques pour chaque entité.

E. Unification des Données de Toutes les Entités

Pour finaliser, nous avons appliqué un processus similaire pour chaque entité et avons fusionné toutes les données en une seule table. Cela a ajouté une nouvelle colonne intitulée "Cubes", représentant les différentes entités.

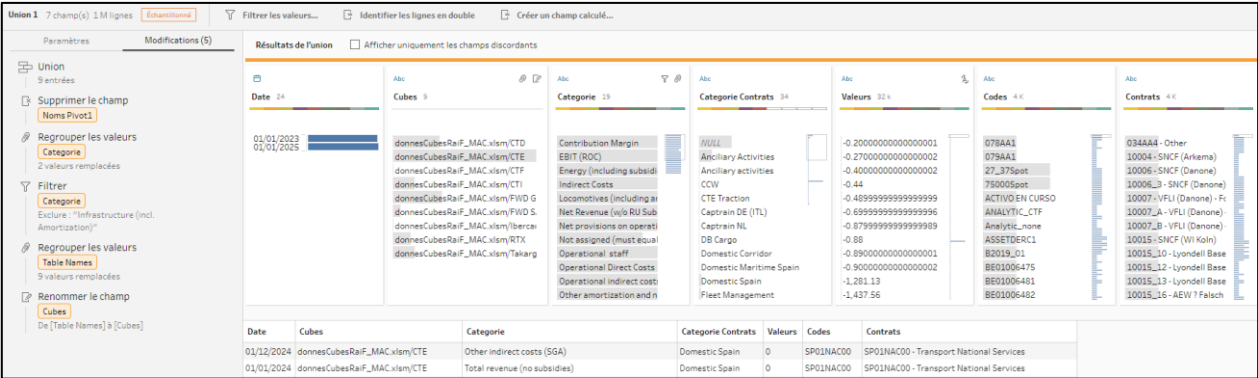


Figure 18 : Données montrant la table unifiée avec la nouvelle colonne "Cubes"

F. Ajustements Finaux avec TabPy

Enfin, nous avons utilisé TabPy, une API intégrée à Tableau pour Python, afin d'effectuer des ajustements finaux sur les données. Cela nous a permis d'affiner la structure de la table, qui contient désormais les colonnes suivantes : Date, Cubes, Catégorie P&L, Catégorie Corridor, Contrats, Code Contrats, et Valeurs.

Date	Cubes	Catégorie	Contrats	Catégorie Contrats	Valeurs	Codes
01/12/2024	donnesCubesRaIF_MAC.xlsm/CTE	Other indirect costs (SGA)	Transport National Services	Domestic Spain	0	SP01NAC00
01/01/2024	donnesCubesRaIF_MAC.xlsm/CTE	Total revenue (no subsidies)	Transport National Services	Domestic Spain	0	SP01NAC00
01/01/2024	donnesCubesRaIF_MAC.xlsm/CTE	RUs Subcontracting costs	Transport National Services	Domestic Spain	0	SP01NAC00
01/01/2024	donnesCubesRaIF_MAC.xlsm/CTE	Net Revenue (w/o RU Subcontracting)	Transport National Services	Domestic Spain	0	SP01NAC00
01/01/2024	donnesCubesRaIF_MAC.xlsm/CTE	Path and tolls (including subsidies)	Transport National Services	Domestic Spain	0	SP01NAC00
01/01/2024	donnesCubesRaIF_MAC.xlsm/CTE	Energy (including subsidies)	Transport National Services	Domestic Spain	0	SP01NAC00
01/01/2024	donnesCubesRaIF_MAC.xlsm/CTE	Locomotives (including amortization)	Transport National Services	Domestic Spain	0	SP01NAC00
01/01/2024	donnesCubesRaIF_MAC.xlsm/CTE	Operational staff	Transport National Services	Domestic Spain	0	SP01NAC00
01/01/2024	donnesCubesRaIF_MAC.xlsm/CTE	Wagons (including amortizations)	Transport National Services	Domestic Spain	0	SP01NAC00
01/01/2024	donnesCubesRaIF_MAC.xlsm/CTE	Other subcontracting services (other than RU) and Pro	Transport National Services	Domestic Spain	0	SP01NAC00
01/01/2024	donnesCubesRaIF_MAC.xlsm/CTE	Net provisions on operational risks	Transport National Services	Domestic Spain	0	SP01NAC00
01/01/2024	donnesCubesRaIF_MAC.xlsm/CTE	Other operational direct costs	Transport National Services	Domestic Spain	0	SP01NAC00

Figure 19 : Table finale prête pour l'analyse

3.3 CHARGEMENT AUTOMATISE DES DONNEES DE BIGQUERY VERS TABLEAU

Pour automatiser le chargement des données de BigQuery vers Tableau, nous avons configuré Tableau Server pour gérer les connexions et les mises à jour de données de manière continue et sans intervention manuelle. Ce processus garantit que les tableaux de bord dans Tableau sont toujours alimentés par les données les plus récentes disponibles dans BigQuery.

Tout d'abord, une connexion en direct a été établie entre Tableau et BigQuery via Tableau Server. Cette connexion permet à Tableau d'accéder directement aux données stockées dans BigQuery, éliminant ainsi le besoin de téléchargements manuels ou de synchronisations périodiques. Ensuite, nous avons configuré Tableau Server pour rafraîchir automatiquement les extraits de données à intervalles réguliers, de façon hebdomadaire, une demande spécifique du projet.

En utilisant Tableau Server, nous avons également mis en place un système de surveillance pour suivre l'état des mises à jour automatiques. Des alertes ont été configurées pour notifier l'équipe en cas d'échec de ces mises à jour, garantissant ainsi que les données disponibles dans les tableaux de bord sont toujours exactes et à jour. Ce processus automatisé a amélioré l'efficacité opérationnelle et a réduit le risque d'erreurs dues à des mises à jour manuelles.

3.4 CREATION DE DASHBOARDS DYNAMIQUES AVEC TABLEAU

Dans cette section, nous avons conçu un tableau de bord dynamique en utilisant Tableau pour visualiser les données financières pertinentes extraites et transformées à partir de BigQuery. L'objectif principal de ce dashboard est de fournir aux utilisateurs, notamment les équipes financières, une vue d'ensemble claire et interactive des performances financières de différentes entités au sein de l'entreprise.

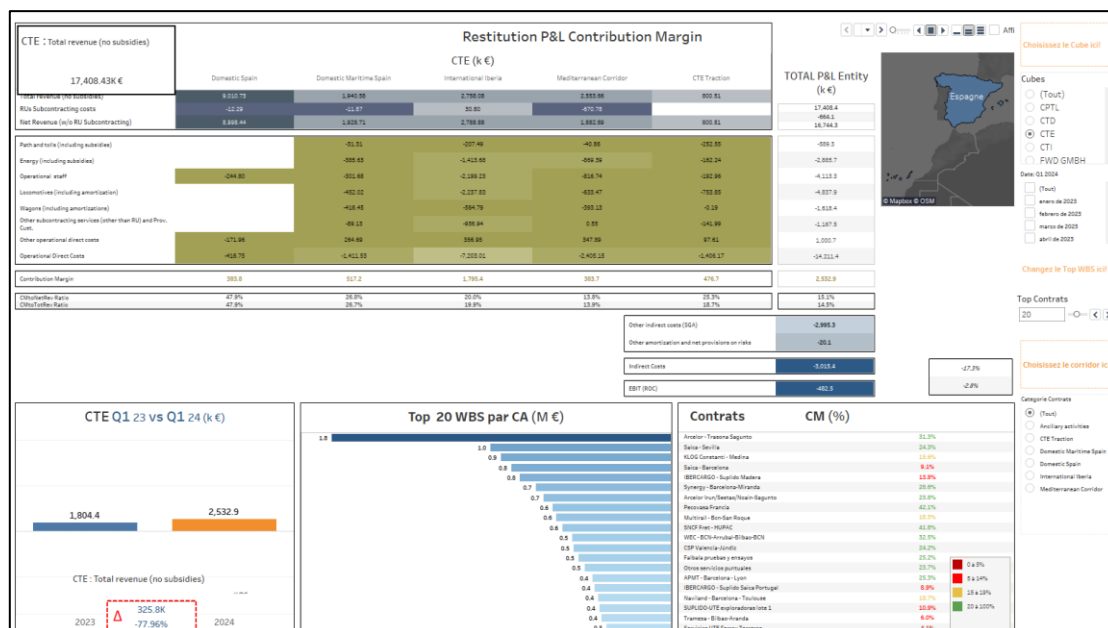


Figure 20 : Résultat final de Dashboard sur Tableau Desktop

Nous pouvons observer, si nous pointons le graphique, que nous avons 5 parties importantes :

- Le tableau dynamique général des P&L.
- La carte géographique indiquant le pays de l'entité,
- La comparaison des valeurs de P&L entre l'année en cours et N-1,
- La liste des 20 premiers contrats avec les Chiffres d'affaires et le taux de marge contribuant au résultat net,
- Les filtres respectifs.

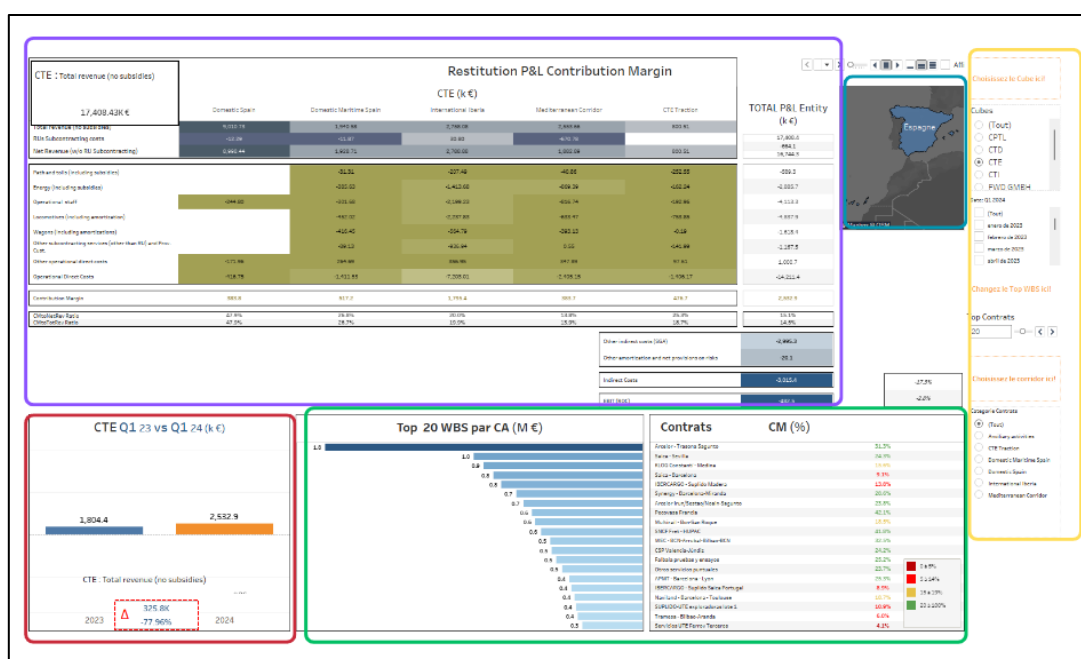


Figure 21 : Détail de 4 parties importantes du graphique

Description du Dashboard:

Vue Globale des Performances Financières: Le dashboard présente une restitution détaillée de la marge contributive (P&L), *dans ce cas pour l'entité CTI*, avec des comparaisons entre différents corridors et segments. La section supérieure montre les revenus totaux sans subventions et d'autres indicateurs clés, segmentés par régions et activités. Cette visualisation permet d'identifier rapidement les secteurs où la performance est positive ou négative, grâce à un code couleur.

Restitution P&L Contribution Margin						
CTI (k €)						
CTI : Total revenue (no subsidies)	Domestic Corridor	Eastern Corridor	Fleet Management	French Corridor	Shunting Contracts	TOTAL P&L Entity (k €)
25,773.92K €						
Total Revenue (no subsidies)	4,818.96	10,126.61	0.28	1,823.78	1,212.04	25,773.9
RUs Subcontracting costs	-380.08	-4,560.16	0.00	-101.34	0.00	-7,681.2
Net Revenue (w/o RU Subcontracting)	4,438.88	5,566.45	0.28	1,722.43	1,212.04	18,092.7
Path and tolls (including subsidies)	-216.02	-321.10	0.00	-89.67	0.00	-851.6
Energy (including subsidies)	-183.37	-144.64	0.00	-39.72	-18.11	-617.2
Operational staff	-1,799.66	-1,981.41	0.00	-599.92	-0.15	-6,329.1
Locomotives (including amortization)	-1,556.01	-1,566.31	-244.89	-337.04	0.00	-4,647.0
Wagons (including amortizations)	-1,155.83	-406.53	0.00	-11.99	0.00	-1,854.4
Other subcontracting services (other than RU) and Prov. C.	-70.81	-102.14	-1.15	-68.76	-100.19	-659.9
Net provisions on operational risks	0.00	0.00	0.00	0.00	0.00	0.0
Other operational direct costs	-13.94	-74.19	-53.72	-6.56	-147.38	-1,193.0
Operational Direct Costs	-4,997.74	-4,596.32	-299.76	-1,153.67	-285.83	-15,752.3
Contribution Margin	-558.9	1,325.0	-299.5	568.8	946.2	2,340.4
CMtoNetRev Ratio	-12.6%	22.4%	-106197.4%	33.0%	78.1%	12.9%
CMtoTotRev Ratio	-11.6%	12.9%	-106197.4%	31.2%	78.1%	9.1%
Operational indirect costs						-17.4
Other indirect costs (SGA)						-1,746.6
Other amortization and net provisions on risks						-140.3
Indirect Costs						-1,904.3
EBIT (ROC)						436.1

Figure 22 : Description général du P&L pour le cas de CTI

Comparaison Temporelle: En bas à gauche, un graphique compare les résultats du premier trimestre 2023 avec ceux du premier trimestre 2024 pour CTE. Les utilisateurs peuvent voir d'un coup d'œil l'évolution des revenus, avec des indicateurs de variation en pourcentage et en valeur absolue, facilitant ainsi l'analyse des tendances.

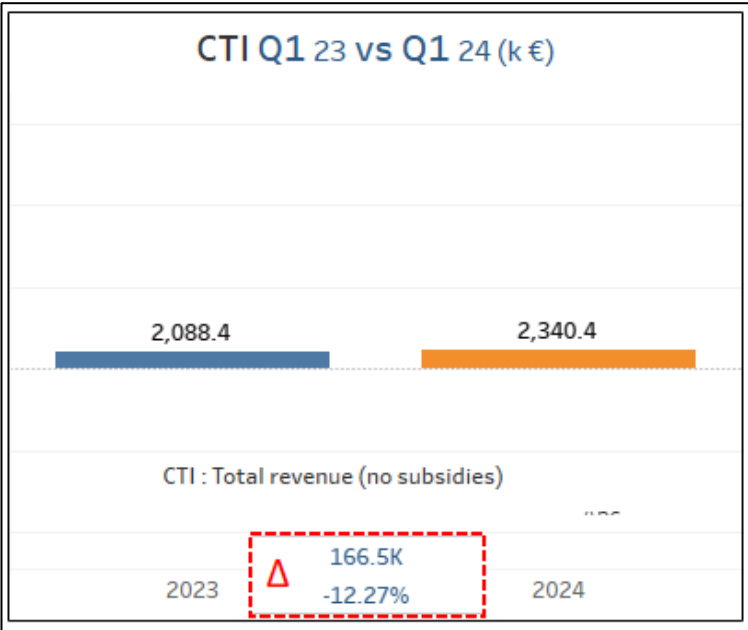


Figure 23 : Comparaison entre les années en cours et l'année N-1, avec écart détaillé en milliers d'euros et pourcentage

Contrats et Contributions: Le dashboard inclut également une section **"Top 20 WBS par CA (M€)"** qui classe les contrats par leur chiffre d'affaires, ainsi qu'une vue détaillée de la marge contributive (CM %) pour ces contrats. Cette section permet aux décideurs de comprendre quels contrats contribuent le plus à la performance globale et lesquels nécessitent une attention particulière.

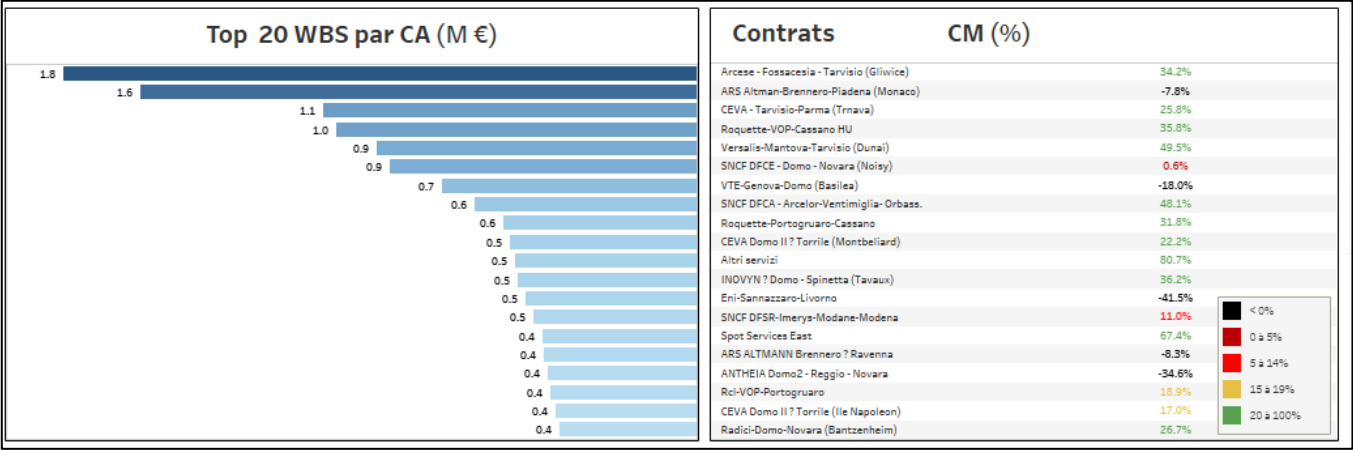


Figure 24 : Top 20 WBS par CA et ratio de CM sur résultat net

Filtrage Interactif: Des filtres interactifs sur le côté droit permettent aux utilisateurs de choisir l'entité ("Cubes"), la période (mois ou trimestre), et le corridor spécifique, leur offrant ainsi la flexibilité d'explorer les données selon différents critères. Cette fonctionnalité assure une analyse personnalisée et adaptée aux besoins spécifiques de chaque utilisateur.

Choisissez le Cube ici!

Cubes

☐ (Tout)

☐ CPTL

☐ CTD

☐ CTE

☒ CTI

☐ FWD GMBH

Date: Q1 2024

☐ (Tout)

☐ enero de 2023

☐ febrero de 2023

☐ marzo de 2023

☐ abril de 2023

Changez le Top WBS ici!

Top Contrats

20

Choisissez le corridor ici!

Categorie Contrats

☒ (Tout)

☐ Domestic Corridor

☐ Eastern Corridor

☐ Fleet Management

☐ French Corridor

☐ North - South Corridor

☐ Other Services

☐ OTHERS

☐ Shunting Contracts

☐ Terminal Services

Figure 25 : Détail de filtres interactifs

Interactivité Dynamique de l'évolution des contrats Top 20 : Les graphiques en bas à droite, affichent les contributions des contrats en termes de marge contributive (CM %) et de chiffre d'affaires (CA) pour les contrats top 20 par CA. Un des aspects les plus puissants de ce dashboard est la possibilité d'interagir avec les données : *en passant la souris sur chaque contrat*, que ce soit sur le graphique des CA des contrats Top 20 ou sur l'évolution en pourcentage du ratio (Contribution Margin/CA net), l'utilisateur peut voir les évolutions précises et détaillées. Cette fonctionnalité permet d'avoir une vue instantanée et approfondie sur l'évolution de chaque contrat au fil du temps, offrant ainsi une meilleure compréhension des performances et des tendances sous-jacentes.

Cette interactivité enrichit l'expérience utilisateur en rendant l'exploration des données plus intuitive et en permettant une analyse plus fine des résultats contractuels. Le tableau de bord devient ainsi un outil essentiel pour les équipes financières, leur fournissant des insights précis et exploitables.

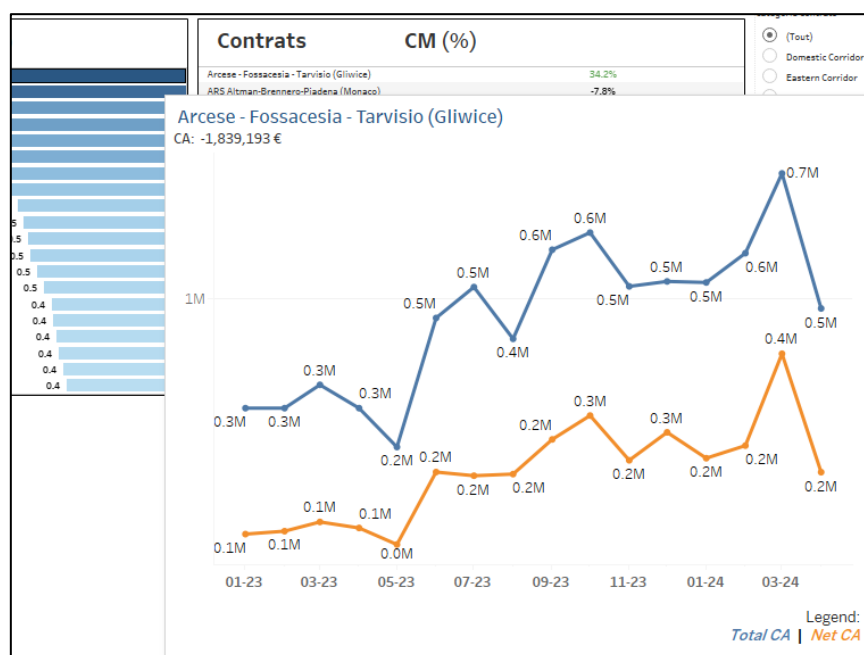


Figure 26 : Evolution du CA et du CA Net pour chaque contrat au survol de souris

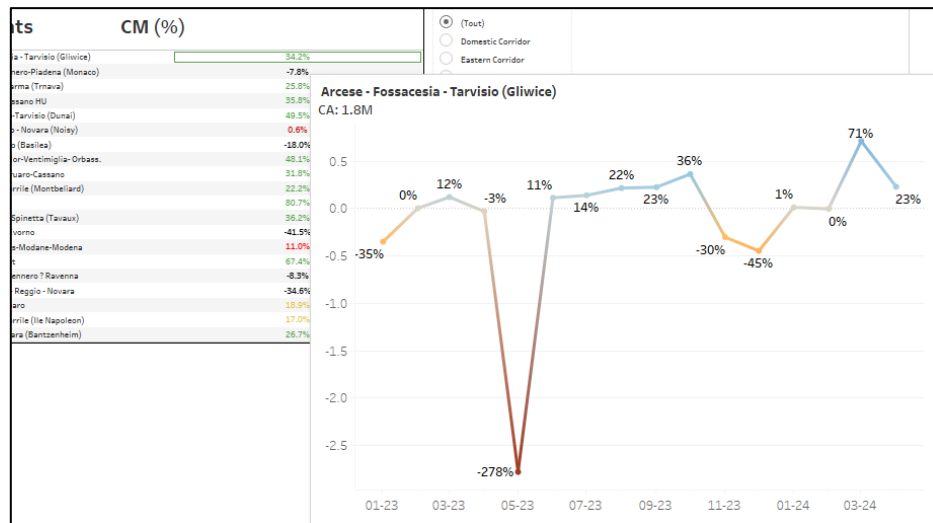


Figure 27 : Evolution du ratio Contribution Margine et CA Net pour chaque contrat au survol de souris

3.5 AUTOMATISATION ET INTÉGRATION CONTINUE

Le processus d'Extract, Transform, Load (ETL) est une méthode d'intégration des données. Comme illustré de manière similaire dans la Figure 6 du chapitre 2, il consiste à extraire des données brutes provenant d'une ou plusieurs sources, à les stocker temporairement dans une zone de staging, à les transformer, puis à les charger dans un espace de stockage à long terme. Ce processus est crucial pour organiser les données en fonction des besoins de la Business Intelligence (BI) et garantir que les données sont propres, cohérentes et fiables pour l'analyse. Les tâches ETL améliorent la qualité en nettoyant les données avant de les transférer vers un autre dépôt, et le respect des meilleures pratiques améliore leur performance et leur évolutivité, résultant en des données plus précises et fiables. [3]

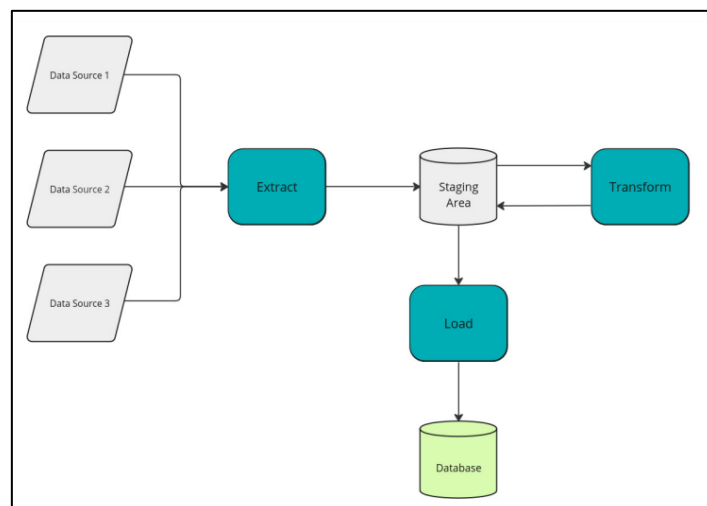


Figure 28 : Exemple de procès ETL [4]

Nous avons mis en place une semi-automatisation du pipeline de traitement des données, optimisant ainsi l'efficacité globale tout en conservant certaines étapes manuelles là où c'était nécessaire.

La première étape du pipeline, réalisée via Excel, reste manuelle. Cependant, grâce à l'utilisation de macros VBA, nous avons considérablement amélioré l'efficacité par rapport à la méthode classique. Ces macros automatisent la collecte et la vérification des données, réduisant ainsi le temps nécessaire pour ces tâches et minimisant les erreurs humaines.

Pour le reste du pipeline, l'automatisation est presque complète. Une fois les données vérifiées et prêtes dans Excel, il suffit de se connecter à Tableau Prep et Tableau Desktop pour que le processus se déroule de manière automatique. Tableau Prep restructure et prépare les données selon les besoins, tandis que Tableau Desktop les visualise directement dans les tableaux de bord.

Cette automatisation partielle mais efficace permet de gagner un temps précieux tout en assurant la fiabilité des données tout au long du processus. De plus, l'intégration continue avec Tableau garantit que les données sont toujours à jour, ce qui facilite la prise de décision rapide et informée par les équipes financières. Cette approche hybride, combinant automatisation et intervention manuelle, nous a permis de maintenir un haut niveau de contrôle tout en profitant des avantages de l'automatisation.

4. ÉVALUATION DE LA SOLUTION ET RESULTATS

L'évaluation de la solution que nous avons mise en place nécessite une approche analytique qui souligne les améliorations obtenues tout en les comparant à l'approche traditionnelle. Cette démarche est essentielle pour démontrer, de manière quantitative et empirique, comment notre solution permet de gagner du temps et d'améliorer l'efficacité des processus décisionnels. En analysant les performances du nouveau pipeline automatisée et en la comparant aux méthodes antérieures, nous pourrions évaluer objectivement l'impact de cette transformation sur l'entreprise.

L'objectif est de montrer concrètement comment l'automatisation, la centralisation des données, et l'amélioration de l'accès à l'information contribuent à une prise de décision plus rapide et plus éclairée. En nous appuyant sur des indicateurs clés de performance (KPI), nous démontrerons les gains de temps, la réduction des erreurs et l'amélioration de la qualité des analyses financières obtenus grâce à cette nouvelle approche.

4.1 PERFORMANCE ET FIABILITE DE LA PIPELINE

Dans notre évaluation de la performance et de la fiabilité du nouveau pipeline automatisée, nous avons pris en compte plusieurs aspects clés, notamment le temps de traitement, la gestion des erreurs, la fréquence des mises à jour, la résilience en cas de panne, et la capacité à effectuer des comparaisons complexes.

Performance et Temps de Traitement

En utilisant le cube "Rail F" sur Oracle ESSBASE, qui contient 9 entités et couvre 7680 contrats pour un total de 15 Mégabytes, nous avons comparé les deux méthodes.

- **Solution classique** : La récupération des données impliquait la création de 9 tableaux distincts, chacun prenant entre 2 et 5 minutes à générer. Le processus complet durait en moyenne entre 30 et 40 minutes, en fonction de la vitesse et de la familiarité des utilisateurs avec les cubes multidimensionnels.
- **Nouvelle pipeline** : Grâce à l'automatisation via les macros VBA, nous avons pu consolider les données en une seule table. Ce processus prend entre 2 et 3 minutes, dépendant uniquement de la vitesse de l'ordinateur. Cela représente une réduction significative du temps de traitement.

Gestion des Erreurs et Vérification des Données

- **Solution classique** : Aucune validation automatique n'était possible. Les équipes devaient vérifier manuellement les données à chaque extraction, ce qui ajoutait quelques minutes supplémentaires à chaque feuille de calcul. Cette méthode augmentait le risque d'erreurs humaines et de données incorrectes.
- **Nouvelle pipeline** : La solution automatisée inclut une étape de vérification dès le lancement de la macro initiale. Cette étape prend environ une minute et génère un rapport qui identifie précisément les données incorrectes, avec des indicateurs visuels (rouge pour les erreurs, vert pour les données correctes). Cela simplifie la notification des erreurs à l'équipe MOA et garantit que toute analyse repose sur des données validées.

Fréquence de Mise à Jour et Résilience

- **Solution classique** : Aucune mise à jour automatique n'était possible, et en cas de panne du système Oracle, l'équipe se retrouvait sans accès aux données.
- **Nouvelle pipeline** : Bien que l'extraction initiale d'Oracle soit toujours manuelle, notre solution offre une mise à jour automatique des données entre BigQuery et Tableau. Cette fonctionnalité assure un accès quasi continu aux données. En cas de panne, les données restent accessibles via BigQuery, assurant ainsi une continuité d'accès et une meilleure résilience.

Comparaisons et Analyses

- **Solution classique** : Les comparaisons complexes entre entités, dates, corridors, contrats, ou pays étaient impraticables sans un travail manuel considérable.
- **Nouvelle pipeline** : Notre solution permet des comparaisons instantanées entre divers éléments, facilitant ainsi les analyses complexes et les prises de décision éclairées.

Suivi et Contrôle des Données

- **Solution classique** : Aucun contrôle numérique n'était disponible, rendant difficile la détection des erreurs ou des modifications non autorisées.
- **Nouvelle pipeline** : Avec l'intégration de Python, nous avons mis en place un suivi des données pour s'assurer qu'elles sont conformes aux données initiales. De plus, tout effacement ou modification accidentelle des données n'affecte que temporairement l'utilisateur, sans compromettre l'intégrité globale des données.

Tableau Comparatif : Solution Classique vs. Pipeline Automatisée Tableau

Critères	Solution Classique (SmartView Excel)	Pipeline Automatisée (Tableau)	Gains de Temps	Réduction (%)
Temps de Traitement	30-40 minutes	2-3 minutes	27-37 minutes	90-93%
Validation des Données	Manuelle (5-10 minutes supplémentaires)	Automatique (1 minute)	4-9 minutes	80-90%
Mise à Jour Automatique	Non disponible	Disponible entre BigQuery et Tableau	Accès continu aux données	N/A
Résilience en Cas de Panne	Aucune	Accès via BigQuery	Accès permanent aux données	N/A
Capacité de Comparaison	Limitée	Complète	N/A	N/A
Suivi des Données	Aucun	Suivi automatisé avec Python	N/A	N/A

4.2 COMPARAISON AVEC L'APPROCHE TRADITIONNELLE (SMARTVIEW EXCEL)

Cette section vise à comparer de manière directe et détaillée la nouvelle solution automatisée avec l'approche traditionnelle utilisant SmartView Excel. En examinant les différences entre ces deux méthodes, nous mettons en lumière les avantages significatifs apportés par l'automatisation, tant en termes de temps que d'efficacité.

Elimination des Goulots d'Étranglement

L'une des principales limitations de l'approche traditionnelle avec SmartView Excel est la nécessité d'une intervention manuelle à chaque étape du processus. Chaque extraction de données doit être effectuée individuellement pour chaque entité, avec des manipulations manuelles pour les assembler en une forme utilisable. Cela crée des goulots d'étranglement, notamment lorsque plusieurs entités ou tableaux doivent être extraits et combinés, ce qui peut prendre un temps considérable et augmenter le risque d'erreurs humaines.

Avec la nouvelle solution automatisée, ces goulots d'étranglement sont pratiquement éliminés. Les macros VBA permettent une extraction et une vérification des données en un temps record, réduisant ainsi le temps total nécessaire pour obtenir une vue d'ensemble des données. De plus, l'automatisation via Tableau Prep et Tableau Desktop permet de consolider et de structurer les données de manière fluide, sans nécessiter de manipulations manuelles complexes.

Facilité d'Utilisation et Intégration

En termes de facilité d'utilisation, l'approche traditionnelle avec SmartView Excel nécessite une bonne connaissance des cubes multidimensionnels et une maîtrise des outils Excel pour naviguer et extraire les données efficacement. Cette méthode peut être chronophage et exige des compétences spécifiques qui ne sont pas toujours présentes au sein de toutes les équipes financières.

À l'inverse, la nouvelle solution offre une interface plus intuitive avec Tableau, facilitant ainsi l'accès aux données même pour les utilisateurs moins expérimentés. L'intégration directe avec BigQuery permet un accès en temps réel aux données centralisées, rendant l'analyse plus rapide et plus accessible. De plus, la connexion automatique entre BigQuery et Tableau garantit que les données sont toujours à jour, éliminant le besoin de rafraîchissements manuels constants.

Comparaison des Temps d'Extraction et d'Analyse

Le temps d'extraction des données est un autre point de comparaison crucial. Avec SmartView Excel, le processus d'extraction pour plusieurs entités peut prendre entre 30 et 40 minutes, en fonction de la complexité des données et de la rapidité de l'utilisateur. En revanche, la solution automatisée réduit ce temps à seulement 2-3 minutes grâce à l'automatisation des étapes d'extraction et de vérification.

4.3 BENEFICES POUR LE PROCESSUS DECISIONNEL FINANCIER

La mise en place de la solution automatisée dans ce projet a apporté des avantages tangibles et significatifs pour le processus décisionnel au sein de l'équipe financière. Elle a, par exemple, renforcé la capacité de l'équipe financière à prendre des décisions éclairées, basées sur des données précises et actualisées. Ces bénéfices se manifestent notamment à travers une plus grande rapidité d'accès aux données, une centralisation efficace de l'information, et des dashboards interactifs qui facilitent les analyses et recommandations stratégiques.

Rapidité d'Accès et Fiabilité des Données

L'un des principaux bénéfices de la solution est la réduction drastique du temps nécessaire pour accéder et traiter les données. Au lieu de passer par des extractions manuelles fastidieuses et potentiellement sources d'erreurs, la solution automatisée permet de centraliser les données dans BigQuery, où elles sont accessibles en temps réel via Tableau. Cette centralisation assure que les équipes financières disposent toujours des informations les plus récentes et fiables pour effectuer leurs analyses.

Grâce à cette rapidité d'accès, les équipes financières peuvent désormais vérifier plus efficacement les allocations de contrats et de WBS dans les corridors appropriés, valider la précision du P&L, et examiner des indicateurs financiers clés tels que l'EBITDA, l'EBIT, le chiffre d'affaires total, et le chiffre d'affaires net. Par exemple, elles peuvent rapidement identifier les contrats à haute marge contributive et formuler des recommandations stratégiques, comme l'optimisation des coûts ou l'ajustement des investissements.

Dashboards Interactifs et Prises de Décision Éclairées

Les dashboards créés avec Tableau offrent une interactivité et une dynamique qui manquaient dans les fichiers Excel traditionnels. Ces visualisations permettent non seulement de présenter les données de manière claire et attrayante, mais aussi de faire des analyses en profondeur sur les performances des contrats et des entités. Par exemple, un directeur financier peut utiliser ces outils pour préparer la lettre de gestion mensuelle en ayant une vue d'ensemble complète des chiffres financiers de toutes les entités. Les informations sont facilement accessibles et peuvent être actualisées en un clic, ce qui est crucial pour la précision et la pertinence des rapports financiers.

De plus, ces dashboards sont utilisés trimestriellement par les entités adjointes et le président de RLE pour évaluer les performances globales et prendre des décisions stratégiques. La solution centralisée ouvre également la voie à une application plus large de cette méthodologie, au-delà de la marge contributive, en intégrant d'autres sujets financiers et opérationnels, ce qui renforce l'efficacité globale du processus décisionnel.

Extension de la Solution à d'autres Sujets

Dans le cadre de ce projet, la solution a été étendue à trois sujets majeurs : le contrôle interne des KPI financiers, les KPI opérationnels des trains et wagons, et la marge contributive des entités. Chaque sujet représentait un défi majeur en termes de traitement des données et de mise en place des analyses, mais la solution a permis de les traiter efficacement et de fournir des insights précieux à l'équipe financière. Cette approche systématique et centralisée s'est révélée être un avantage considérable pour l'équipe, renforçant leur capacité à gérer et analyser des volumes de données complexes de manière cohérente et structurée.

4.4 RETOUR D'EXPERIENCE DES UTILISATEURS FINANCIERS

La mise en place de la nouvelle solution a suscité un retour d'expérience très positif de la part des utilisateurs finaux, en particulier des membres de l'équipe financière. Bien que l'adaptation à cette nouvelle approche ait nécessité un certain temps, les avantages perçus ont rapidement dépassé les défis initiaux, et l'équipe a rapidement adopté la solution comme un atout majeur dans leur travail quotidien.

Feedback des Utilisateurs

Guillaume H., Manager de l'équipe de Contrôle de Gestion :

« J'ai été très impressionné par la manière dont la solution a amélioré notre flux de travail. Les temps de traitement des données ont été drastiquement réduits, et nous avons maintenant accès à des données plus fiables et mieux organisées. Le travail de Ray a vraiment été un atout pour notre équipe, en nous permettant de nous concentrer davantage sur l'analyse plutôt que sur les tâches répétitives de collecte de données. »

Leila T., Collaboratrice en Contrôle de Gestion, en charge de la Lettre de Gestion :

« Grâce à la centralisation des données et aux dashboards interactifs, j'ai pu rédiger la lettre de gestion beaucoup plus efficacement. Auparavant, il fallait rassembler et vérifier les données manuellement, ce qui prenait beaucoup de temps. Maintenant, je peux accéder à toutes les informations dont j'ai besoin en un clic, ce qui me permet de me concentrer sur l'analyse des chiffres. »

Emanuelle E., Collaborateur en Contrôle de Gestion au sein de la Holding RLE :

« L'outil que Ray a mis en place a rendu notre travail beaucoup plus fluide. Nous n'avons plus à nous inquiéter des pannes de système ou des données manquantes. La solution nous offre une fiabilité que nous n'avions pas auparavant, et cela a vraiment amélioré notre capacité à prendre des décisions stratégiques. »

Karilm B., Contrôleur de Gestion Senior :

« En tant que contrôleur de gestion senior, je dois superviser un grand volume de données. Avant, c'était un vrai défi d'assurer la cohérence et l'exactitude des informations entre les différents tableaux Excel. Avec cette nouvelle solution, les vérifications sont automatisées, et je peux enfin me concentrer sur les recommandations stratégiques au lieu de perdre du temps sur des tâches répétitives. »

Adaptation et Collaboration

Au départ, l'introduction de nouveaux outils et d'une nouvelle méthodologie a suscité des défis d'adaptation, tant pour moi, en tant que Data Analyst/Scientist, que pour l'équipe financière. La transition a nécessité un travail de vulgarisation des concepts techniques, tels que le Big Data, l'intelligence artificielle, et les décisions basées sur les données (data-driven decisions), pour assurer une compréhension mutuelle et une collaboration efficace.

Cependant, malgré ces défis initiaux, l'équipe a progressivement adopté la solution. Grâce à une communication continue et à des séances de formation ciblées, nous avons réussi à intégrer cette nouvelle approche dans les pratiques quotidiennes, apportant une réelle valeur ajoutée à l'entreprise. Cette expérience a non seulement renforcé l'efficacité de l'équipe, mais elle a également démontré l'importance de l'intégration des technologies avancées dans les processus financiers.

5. MISE EN PLACE DU PROJET «SOLUTION »

La solution et sa mise en œuvre sont deux étapes séquentielles dans le développement de notre mémoire de recherche. Une fois que nous avons planifié les hypothèses de travail et d'utilisation, il est nécessaire d'établir une liste des tâches à accomplir, de définir les jalons, les périodes, ainsi que de prévoir un budget de travail, incluant les coûts des logiciels. Cependant, la partie budgétaire du projet sera abordée dans le chapitre 7 de ce mémoire.

5.1 FINALITE DU PROJET

La finalité de ce projet est de créer un pipeline de données automatisée qui améliore la gestion et l'analyse des données financières au sein du service financier de l'entreprise. L'objectif est de surmonter les limitations actuelles liées à l'utilisation de SmartView Excel pour l'extraction des données d'Oracle Essbase, en offrant une solution plus agile, fiable, et interactive grâce à l'intégration de BigQuery et Tableau. Cette solution vise non seulement à accélérer les processus décisionnels, mais aussi à augmenter la précision des analyses et à offrir un accès continu aux données critiques, même en cas de défaillance des systèmes principaux.

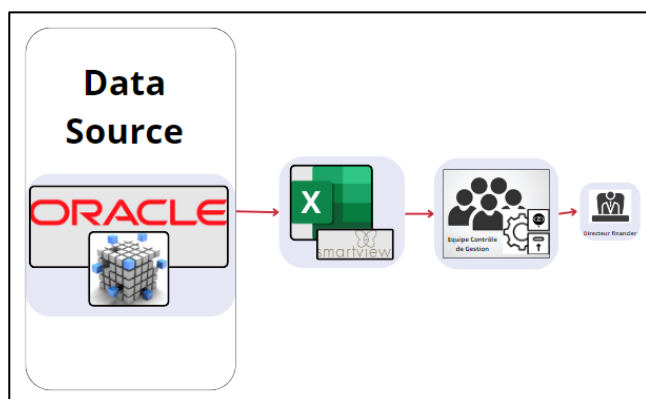


Figure 29 : Procès de travail classique

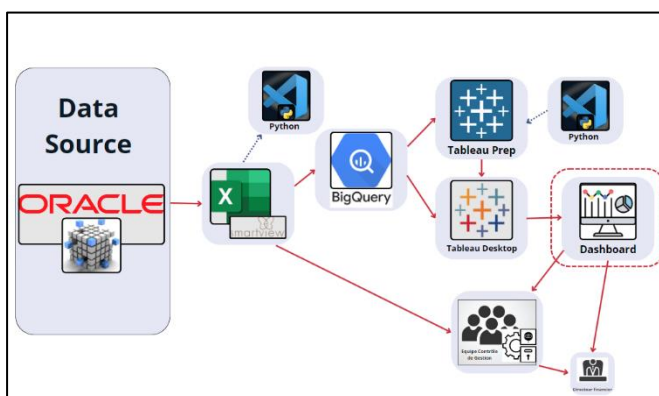


Figure 30 : Process avec un pipeline intermédiaire

Nous avons rapidement remarqué la redondance des données et l'ajout de contrôles supplémentaires dans le processus classique sans perturber l'équipe financière ni le directeur financier, qui sont les principaux utilisateurs de ces données. Ces ajustements ont permis d'assurer une meilleure fiabilité des informations tout en offrant une continuité de l'accès aux données critiques, essentielle pour les prises de décision stratégiques au sein de l'entreprise.

5.2 LISTE DES TACHES A EFFECTUER

Pour atteindre les objectifs fixés, les tâches suivantes ont été identifiées et mises en œuvre en utilisant des informations de planification des projets pertinents [\[1\]](#) [\[2\]](#) :

1. **Analyse des besoins** : Identifier les besoins spécifiques de l'équipe financière en matière de données.
2. **Mise en place des macros VBA** : Automatiser l'extraction des données depuis SmartView Excel.
3. **Configuration de BigQuery** : Créer un Data Warehouse pour centraliser les données extraites.
4. **Développement des transformations dans Tableau Prep** : Structurer et nettoyer les données pour qu'elles soient prêtes à être visualisées.
5. **Création des dashboards dans Tableau Desktop** : Développer des visualisations dynamiques pour faciliter l'analyse.
6. **Automatisation du pipeline** : Assurer une mise à jour régulière et automatique des données et des dashboards.
7. **Tests et ajustements** : Valider la fiabilité de la solution et apporter les ajustements nécessaires.
8. **Documentation et formation des utilisateurs** : Rédiger la documentation du projet et former les utilisateurs finaux à l'utilisation des nouveaux outils.

5.3 DUREE DU PROJET, PLANIFICATION ET JALONS

Durée du Projet

Le projet a été réalisé sur une période de huit mois, allant de janvier à août 2024, avec une intensification des efforts à partir de mai 2024. La phase initiale du projet, de janvier à avril, a été consacrée à l'analyse des besoins, à la mise en place des outils (VBA pour Excel, BigQuery), et au développement initial du pipeline de données. Pendant cette période, le temps de travail était réparti sur trois jours par semaine. La phase suivante, de mai à juillet, a vu une augmentation du rythme avec cinq jours par semaine dédiés au projet. Cette phase a été marquée par des ajustements finaux, des tests rigoureux, la documentation du projet, et la préparation des livrables. Le projet s'est conclu en août avec la finalisation de la mise en production et la formation des utilisateurs finaux, garantissant une transition en douceur vers l'utilisation de la nouvelle solution.

Planification et Jalons

Le projet a commencé en janvier 2024 et s'est terminé en juillet 2024. La planification a été divisée en deux phases principales, avec un calendrier adapté aux disponibilités et aux priorités du projet :

Phase 1 : Janvier à Mai 2024 (3 jours par semaine)

- **Janvier** : Analyse des besoins et collecte des données initiales depuis Oracle Essbase. Mise en place des macros VBA pour l'automatisation de l'extraction.
- **Février** : Transfert des données dans BigQuery et mise en place des premières tables. Début du développement des transformations dans Tableau Prep.
- **Mars** : Test et optimisation du pipeline entre BigQuery et Tableau Prep. Développement initial des dashboards dans Tableau Desktop.
- **Avril** : Finalisation des transformations dans Tableau Prep et des calculs avancés. Début de l'automatisation des mises à jour.
- **Mai** : Ajustements et tests finaux du pipeline. Préparation des données pour la phase d'utilisation intensive.

Phase 2 : Mai à Juillet 2024 (5 jours par semaine)

- **Mai** : Intensification du travail sur les dashboards Tableau, avec un focus sur la création d'insights dynamiques et l'intégration de KPIs spécifiques.
- **Juin** : Tests finaux et validation complète du pipeline automatisée. Documentation du projet et préparation des livrables.
- **Juillet** : Finalisation du projet, revue des performances du pipeline, et mise en production. Présentation des résultats et formation des utilisateurs finaux.

Le schéma de travail a été élaboré à l'aide de l'outil Trello, une plateforme particulièrement efficace pour le suivi de tout type de projet. Trello ne se limite pas à la simple planification des tâches à accomplir ; il permet également d'enregistrer et de gérer des images, vidéos, liens, référentiels GitHub, et bien plus encore. Ces fonctionnalités ont motivé mon choix de cet outil pour la gestion de ce mémoire de recherche. Ci-dessous, vous trouverez un aperçu des tâches

déjà réalisées dans Trello, qui illustre le déroulement du travail, les jalons atteints, les livrables, ainsi que les périodes spécifiques en fonction de ma disponibilité en tant qu'alternant.



Figure 31 : Planification et organisation sur Trello

6. PERSPECTIVES D'AMÉLIORATION ET EXTENSION

Dans cette section, nous plongeons dans la phase de recherche interne et externe de besoins, pour proposer des solutions potentielles pour améliorer le projet et mieux répondre à notre problématique.

6.1 OPTIMISATION CONTINUE ET SCALABILITÉ

Avec une mise à jour hebdomadaire des outils comme Tableau et BigQuery, le pipeline de données est déjà bien optimisé pour les besoins actuels de l'entreprise. Cependant, des opportunités d'amélioration existent, en particulier concernant l'extraction des données depuis Oracle Essbase. Ce processus reste un point de goulot d'étranglement, nécessitant une optimisation pour réduire le temps d'extraction et maximiser l'efficacité. Une solution pourrait être l'intégration de nouveaux outils ou scripts pour automatiser davantage cette étape ou même envisager une migration des calculs de Oracle vers BigQuery pour centraliser les opérations et gagner en rapidité.

À mesure que l'entreprise prévoit une croissance significative des volumes de données, notamment avec l'intégration de données opérationnelles massives de la SNCF, il devient essentiel de préparer la solution pour une montée en échelle. Le passage des centaines de gigaoctets aux dizaines de pétaoctets nécessitera des ajustements au niveau de l'architecture des données pour garantir que le pipeline reste performant et capable de traiter de grands volumes de données sans compromettre la qualité ou la rapidité.

6.2 INTEGRATION DE NOUVELLES SOURCES DE DONNEES

L'intégration de nouvelles sources de données, telles que SAP4HANA et les bases de données associées à Salesforce, pourrait transformer l'analyse financière en offrant une vue plus complète et plus rapide des données. SAP4HANA, par exemple, fournirait une voie de sécurité plus efficace, mais nécessiterait une adaptation car elle ne fournit pas les calculs financiers déjà disponibles dans Oracle Essbase. Le défi technique ici réside dans la migration des calculs vers BigQuery et la gestion des accès et droits, surtout dans un environnement où la sécurité et la confidentialité des données financières sont cruciales.

En outre, convaincre toutes les entités européennes d'adopter un système unifié pourrait représenter un défi organisationnel majeur. Cependant, une fois intégré, un tel système offrirait une cohérence et une efficacité accrues dans l'analyse des données financières, rendant la prise de décision plus rapide et plus fiable.

6.3 AUTOMATISATION AVANCEE AVEC MACHINE LEARNING

L'intégration de l'intelligence artificielle dans le pipeline pourrait automatiser plusieurs aspects critiques de l'analyse des données. Par exemple, des modèles de machine learning pourraient être utilisés pour classer automatiquement les contrats dans les corridors financiers, une tâche actuellement effectuée manuellement. De plus, la prévision de KPI financiers, tels que l'EBITDA, le ROI, et le chiffre d'affaires, pourrait être automatisée, offrant des prédictions plus précises basées sur l'ensemble des données disponibles.

Le développement de modèles prédictifs pour identifier les entités nécessitant une attention particulière dans les mois à venir serait un autre exemple de l'application du machine learning. Cela permettrait à l'équipe de contrôle de gestion de se concentrer sur les aspects financiers critiques tout en optimisant leur charge de travail.

À long terme, l'objectif est de faciliter le travail de l'équipe de contrôle de gestion, leur permettant de se concentrer sur l'analyse financière et la construction de rapports, tout en laissant l'intelligence artificielle s'occuper des tâches de prévision et de classification. Cela améliorerait non seulement l'efficacité, mais aussi la qualité des décisions financières, renforçant ainsi la compétitivité de l'entreprise dans un environnement économique de plus en plus complexe.

7. REALISATION D'UN BUSINESS PLAN

Dans cette section, nous entreprenons la création d'un Business Plan. Notre objectif est d'évaluer l'impact financier de notre projet, de considérer sa viabilité à long terme et d'explorer les implications pratiques pour la gestion et la prise de décision financière.

Estimation des Coûts

Dans cette phase, nous identifierons et évaluerons les coûts associés à la mise en œuvre du projet. Cela inclut les coûts liés aux frais de développement logiciel, les coûts liés à l'utilisation d'outils d'analyse et de visualisation.

Catégorie de Coûts	Description	Montant Estimé (euros)
Logiciels		
- Tableau (Licences)	1 licence créateur (1380€/an), 2 licences viewer (840€/an chacune), 1 licence explorer (840€/an)	€ 3,060.00
- BigQuery (Requêtes)	Coût pour le traitement des requêtes SQL (estimé sur usage mensuel)	€ 500.00
- Oracle (Licence Utilisateur)	Licence utilisateur pour l'accès aux données sur Oracle Essbase	€ 479.00
- Python	Gratuit	€ -
Main-d'Œuvre		
- Data Analyst Junior	Salaire pour 7 mois (Janvier à Juillet, 3 jours/sem. jusqu'à mai, 5 jours/sem. ensuite)	€ 12,000.00
- Manager (temps passé sur le projet)	Estimation du temps de supervision et de validation	€ 5,000.00
Maintenance et Support		
- Mise à jour des logiciels	Coût des mises à jour annuelles pour Tableau, BigQuery, et autres logiciels	€ 800.00
- Support Technique	Assistance technique pour l'installation et la gestion des logiciels	€ 1,000.00
Autres Coûts		
- Électricité et Frais Divers	Estimation des coûts énergétiques et autres frais liés au projet	€ 200.00
Total des Coûts		€ 23,039.00

7.1 ÉVALUATION FINANCIERE DU PROJET

Coûts des Licences Logicielles :

Le projet utilise plusieurs outils essentiels, dont Tableau et BigQuery, pour la gestion et l'analyse des données. Le coût total des licences pour Tableau est estimé à 3 060 euros par an. Ce montant comprend une licence « créateur » à 1 380 euros par an, deux licences « viewer » à 840 euros chacune, et une licence « explorer » à 840 euros. Le coût de BigQuery dépend principalement de l'utilisation, avec des tarifs de 6,25 USD par TiB pour les requêtes, bien que la première portion d'utilisation soit gratuite chaque mois. La tarification flexible de BigQuery est avantageuse car elle permet de payer uniquement pour les requêtes effectivement utilisées, ce qui peut être économique étant donné le nombre limité de requêtes mensuelles. En outre, la licence utilisateur pour Oracle coûte 479 euros par utilisateur, ce qui ajoute aux coûts globaux du projet. Python, utilisé pour certaines étapes du processus, est gratuit, ce qui aide à réduire les coûts de développement.

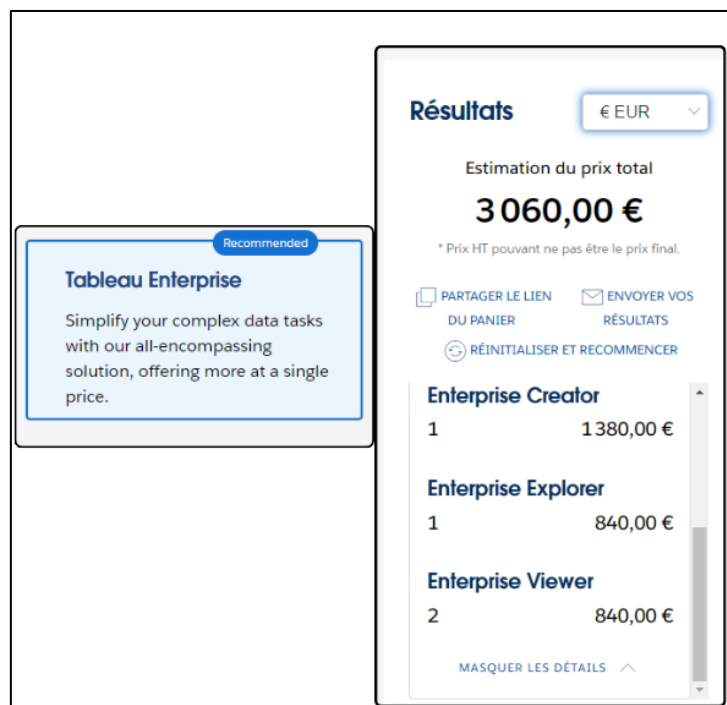


Figure 32 : Exemple de coûts de licence Tableau Software [11]

Temps et Ressources Humaines :

La mise en œuvre du projet a été réalisée par un data analyste junior, en collaboration avec un manager de l'équipe de contrôle de gestion, qui a agi en tant que client pour valider les solutions proposées. Le projet a débuté en janvier et s'est poursuivi jusqu'en juillet, avec un engagement de 3 jours par semaine jusqu'à mai, puis de 5 jours par semaine jusqu'à sa finalisation en juillet.

Frais de Maintenance :

Les coûts récurrents de maintenance incluent les mises à jour logicielles, notamment pour Tableau. Il est essentiel de maintenir à jour ces outils pour garantir la sécurité et la performance du pipeline. En général, les coûts de maintenance des serveurs et de support technique doivent également être pris en compte. Selon certaines sources, les coûts de maintenance pour les solutions de BI peuvent représenter environ 15 % des coûts initiaux d'installation.

Bénéfices Attendues :

L'automatisation du pipeline a permis de gagner entre 80 et 90 % de temps, ce qui représente un gain immédiat en termes de productivité. Ce gain pourrait encore s'améliorer avec le temps, à mesure que l'équipe s'habitue aux nouvelles technologies et que d'éventuelles optimisations sont mises en œuvre.

Réduction des Erreurs :

L'automatisation a permis de réduire les erreurs de 99 %, ce qui a considérablement amélioré la qualité des données et, par conséquent, la prise de décision. Cette réduction drastique des erreurs se traduit non seulement par une meilleure fiabilité des données, mais également par une amélioration notable dans l'efficacité des processus décisionnels au sein de l'entreprise.

7.2 PERSPECTIVES DE GESTION ET ANALYSES FINANCIERES

Ci-dessous un tableau qui nous aide à avoir une vision globale des perspectives financières du projet :

Aspect	Description	Impact Potentiel	Difficultés Anticipées	Actions Correctives Proposées
Adaptabilité de la Solution	La solution peut être étendue à d'autres départements (Ressources Humaines, Comptabilité, etc.).	Facilite l'intégration des données dans différents secteurs, optimisant les processus de décision.	Résistance au changement et manque de compétence dans d'autres départements.	Formation continue et support technique dédié pour faciliter l'adoption.
Intégration Future d'Outils	Possibilité d'intégrer d'autres outils analytiques ou technologiques (par ex. SAP4hana).	Améliore l'efficacité et la sécurité des opérations tout en facilitant l'analyse financière globale.	Coût supplémentaire et complexité d'intégration technique.	Planification budgétaire et essais pilotes pour minimiser les risques avant l'intégration
Stratégie Financière à Long Terme	Contribue aux objectifs globaux de digitalisation et d'optimisation des coûts.	Renforce la stratégie financière de l'entreprise, améliorant la compétitivité et la rentabilité.	Nécessité de justifier les investissements à long terme auprès des parties prenantes.	Communication régulière des bénéfices réalisés et projections à long terme pour convaincre les
Scalabilité	La solution est conçue pour évoluer avec l'augmentation des volumes de données et	Réduit les risques de surcharge, assurant la continuité et la performance du système.	Problèmes potentiels de performance avec l'augmentation rapide des	Suivi et ajustements constants des ressources allouées pour garantir la
Gestion des Risques	Identification et gestion des risques techniques, financiers, et humains.	Assure une mise en œuvre fluide et une adoption efficace de la solution par les équipes.	Risques de sous-estimation de l'impact des risques identifiés.	Évaluation continue des risques et mise à jour des plans de gestion des risques.

Figure 33 : Résumé des principales perspectives de gestion et d'analyse financière du projet

Adaptabilité de la Solution :

La solution actuelle pourrait facilement être étendue à d'autres départements, tels que les ressources humaines et la comptabilité, qui utilisent également Oracle Essbase. De plus, tous les départements de contrôle financier en Europe pourraient bénéficier de cette solution centralisée, augmentant ainsi l'efficacité opérationnelle à un niveau transnational.

Intégration Future d'Outils :

Bien que des outils spécifiques ne soient pas encore identifiés pour une intégration future, des options telles que **Alteryx** pour l'automatisation des processus analytiques, **Power BI** pour des fonctionnalités complémentaires de visualisation, ou même **SAP Analytics Cloud** pour une intégration plus étroite avec les systèmes SAP, pourraient être envisagées. Ces outils pourraient enrichir les capacités analytiques et permettre une gestion encore plus précise et réactive des données.

Stratégie Financière à Long Terme :

Bien que la solution puisse sembler coûteuse au départ, les résultats attendus, tels que des gains de productivité et une meilleure précision des données, justifient pleinement cet investissement. À long terme, cette solution s'intègre dans la stratégie financière globale de l'entreprise, contribuant à des objectifs plus larges comme la digitalisation des processus et l'optimisation des coûts opérationnels. Cette digitalisation renforce l'agilité de l'entreprise, la préparant mieux à répondre aux défis futurs du marché.

Scalabilité :

La solution est conçue pour être scalable, ce qui signifie qu'elle peut évoluer facilement en fonction de l'augmentation des volumes de données ou du nombre d'utilisateurs. Cependant, cela nécessitera des ajustements tels que l'ajout de capacités de stockage ou l'optimisation des requêtes pour maintenir la performance du système à mesure que les besoins augmentent.

Risques :

Les principaux risques incluent des défis techniques liés à la mise en œuvre, en particulier avec les services informatiques pour l'installation de logiciels et de pilotes. Sur le plan financier, bien que le budget soit disponible pour une mise en œuvre progressive à partir de 2025, il est crucial de gérer ces ressources avec prudence. Enfin, sur le plan humain, il sera essentiel de former le personnel pour qu'il s'adapte aux nouvelles technologies, en surmontant les réticences liées au changement de méthodes de travail.

8. CONCLUSIONS ET PERSPECTIVES FUTURE

En conclusion de mon mémoire de recherche, il est évident que ce projet m'a apporté bien plus que des compétences techniques. Il m'a permis de développer des compétences transversales (soft skills) et d'acquérir une expertise approfondie en data science, particulièrement dans le domaine de la visualisation de données (Data Viz) et de la création de pipelines adaptés à des problématiques spécifiques. Grâce à ce projet en entreprise, j'ai eu l'opportunité de travailler non seulement avec Tableau pour générer des insights, mais aussi avec d'autres outils tels que Power BI, Looker Studio, et Excel avancé. Ces expériences variées m'ont permis de maîtriser divers outils de Business Intelligence, renforçant ainsi mon affinité professionnelle pour ce domaine que je continue à explorer et à approfondir.

8.1 SYNTHESE DES RESULTATS ET CONCLUSIONS

Le projet a permis de mettre en place une solution innovante et efficace pour la gestion des données financières au sein de l'entreprise. Les principaux défis rencontrés, notamment la gestion du temps et les obstacles techniques liés à l'intégration des outils, ont été surmontés grâce à un travail rigoureux de recherche et de planification. **Grâce à l'automatisation des processus financiers et à l'utilisation d'outils comme Tableau et BigQuery, nous avons réussi à réduire le temps de traitement des données de 85 à 90 %,** ce qui a considérablement amélioré la productivité et la précision des analyses. Les retours des utilisateurs ont été largement positifs, confirmant que la solution répond aux besoins spécifiques du département financier tout en offrant une flexibilité pour des développements futurs.

8.2 CONTRIBUTION DU PROJET A L'ENTREPRISE

Ce projet a eu un impact significatif sur l'efficacité opérationnelle de l'entreprise. **En plus de la réduction du temps nécessaire à l'extraction et à l'analyse des données,** la solution a permis d'augmenter la réactivité des équipes financières face aux défis du marché. La centralisation des données dans BigQuery et l'utilisation de Tableau pour la visualisation ont non seulement amélioré la qualité des analyses, mais ont également renforcé la capacité de l'entreprise à prendre des décisions éclairées. **De plus, la promotion de l'utilisation de l'IA et des outils BI a été essentielle pour intégrer une culture de la donnée au sein de l'équipe, permettant une meilleure adoption des nouvelles technologies.** La solution a posé les bases pour une

meilleure gestion des données à long terme, ce qui est essentiel pour soutenir les objectifs stratégiques de l'entreprise, notamment en matière de digitalisation et d'optimisation des coûts.

8.3 PERSPECTIVES DE DÉVELOPPEMENT FUTUR

Le succès de ce projet ouvre la voie à plusieurs opportunités de développement futur. Premièrement, l'intégration de nouvelles sources de données, telles que les systèmes CRM ou les données de marché externes, pourrait enrichir les analyses financières et offrir des perspectives encore plus larges. **Deuxièmement, l'approfondissement de l'utilisation de l'IA et du machine learning, déjà promus dans ce projet, pourrait permettre de développer des modèles prédictifs pour automatiser encore davantage l'analyse des données et anticiper les tendances financières.** Enfin, la solution actuelle pourrait être adaptée et étendue à d'autres départements, tels que les ressources humaines ou la comptabilité, permettant ainsi à l'entreprise de tirer parti des avantages de l'automatisation à travers l'ensemble de ses activités.

9. BIBLIOGRAPHIE

Cette section présente une liste des sources bibliographiques consultées et référencées tout au long de ce travail de recherche. Ces sources ont été utilisées pour soutenir l'analyse, les méthodologies, les résultats et les conclusions présentés dans ce document. Veuillez consulter ces références pour obtenir plus d'informations sur les concepts abordés et pour approfondir vos connaissances dans le domaine du pipeline de Données Automatisée : Extraction, Transformation et Visualisation via Tableau et des méthodes de visualisation des données.

Livres:

1. Car Allchin. (2020). *Tableau Prep: Up & Running: Self-Service Data Preparation for Better Analysis*
2. Lorna Brown (2021) *Tableau Desktop Cookbook: Quick & Simple Recipes to Help You Navigate Tableau Desktop*
3. Ryan Sleeper (2020) *Innovative Tableau: 100 More Tips, Tutorials, and Strategies*
4. Andy Kirk (2019) *Data Visualisation: A Handbook for Data Driven Design, 2nd Edition*

Articles de recherche et Sites Web:

[1] *Project Management Best Practices: Planning the Project* (2012) "Practice #5: Write a plan" Retrieved from <https://www.projecttimes.com/articles/project-management-best-practices-planning-the-project/>

[2] *All you need to know about project management best practices* (2021) "The 9 best practices for your project" Retrieved from <https://monday.com/blog/project-management/project-management-best-practices/>

[3]. ŠURMAN (2024) "S. Automation of ETL process and data visualization" "Background and Literature Review; 2.1 ETL Process

[4] ŠURMAN (2024) "S. Automation of ETL process and data visualization" "Background and Literature Review; "Figure 2.1: A graphical illustration of ETL process"

[5] DafMag 2019. “Les Deeptech au secours du Contrôle de Gestion” Retrieved from <https://www.daf-mag.fr/Thematique/fonction-finance-1242/Breves/Les-Deeptech-secours-Controle-Gestion-queelles-applications-Machine-Learning-RPA-342807.htm>

[6] Samuel Sponem (2018), “Une “ société du contrôle ” sans contrôle de gestion ? Réflexions sur le Big Data”, La figure 1 «Domaines de compétence des contrôleurs et des data scientist »

[7] Sabrina Lynda ZAIDI (2020), « L’influence du «Big Data » sur le contrôle de Gestion »

[8] BPayd, 2023, « DSO, DPO et DIO : les 3 indicateurs de trésorerie à maîtriser. » Retrieved from <https://bpayd.fr/blog/dso-dpo-dio-definition-calcul-interpretation/>

[9] DafMag 2022. « L'IA dans la finance, une réalité pour Criteo » <https://www.daf-mag.fr/Thematique/BI-1244/outils-si-2132/Breves/L-IA-dans-la-finance-une-realite-pour-Criteo-371683.htm>

[10] Docs Oracle, 2022, «Création d'une connexion à un serveur Essbase » Retrieved from https://docs.oracle.com/applications/help/fr/enterprise-performance-management/11.2/EPMSA/creating_essbase_server_connection.htm#EPMSA-epm_prov_flow_421

[11] Tableau Software, 2024, “Product and pricing selector” Retrieved from <https://www.tableau.com/fr-fr/product-and-pricing-selector>

[12] Google Cloud, 2024. “Tarifs de BigQuery” Retrieved from https://cloud.google.com/bigquery/pricing?gad_source=1&gclid=CjwKCAjwodC2BhAHEiwAE67hJPXQfhcX7Ne52zsPS671FzZakQhrShg89oAEL6cOlRnKykNME25yFBoC298QAvD_BwE&gclsrc=aw.ds&hl=fr

[13] Logiciels pro : Oracle logiciels, 2024, «Oracle Essbase : Avis, Prix, Présentation et Alternatives »

<https://www.logiciels.pro/logiciel-saas/oracle-essbase/>

[14]. "lite finance for trade information" pour comparer les prévisions futures

<https://my.litefinance.org/fr/trading/chart?symbol=XAUUSD&returnUrl=true>

[15]. Johnson, A. B., & Williams, C. D. (2019). "Comparative Analysis of Gold Price Prediction Models." *Journal of Financial Analytics*, 5(2), 45-62.

10.ANNEXES

Dans cette section, vous trouverez des éléments complémentaires qui fournissent une perspective plus approfondie sur le travail de recherche présenté, ainsi que des ressources pratiques pour reproduire ou étendre la solution développée.

Annexe A : Code Source sur GitHub

Le code source utilisé pour l'automatisation du pipeline de données, incluant les scripts Python pour la gestion des données et les codes VBA, est disponible sur GitHub sous le nom de « cm_data.ipynb ». Vous pouvez le consulter et le cloner à l'adresse suivante :

https://github.com/Leonardo-VERA/M2_projet_etude/blob/main/cm_data.ipynb

Annexe B : Données Financières pour l'Étude

Les données utilisées pour cette étude, notamment les fichiers CSV extraits d'Oracle Essbase, sont en partie confidentielles. Cependant, un exemple de données pour l'entité CTI, avec une partie des données aléatorisées via Python (données non réelles mais adéquates pour l'étude), est disponible sur GitHub sous le nom de « CTI.csv ». Ce fichier contient les données de marge contributive pour l'entité CTI. Vous pouvez accéder à ce fichier via le lien suivant :

https://github.com/Leonardo-VERA/M2_projet_etude/blob/main/CTI.csv

Annexe C : Documentation Technique

Une documentation détaillée sur la mise en place du pipeline automatisé, incluant les instructions pour l'installation des outils (Tableau, BigQuery, Python), les codes utilisés dans Tableau Prep, ainsi que les étapes pour reproduire les analyses et les dashboards, est également disponible sur le dépôt GitHub. Cette documentation se trouve dans le fichier « README.md » du même projet GitHub :

https://github.com/Leonardo-VERA/M2_projet_etude/blob/main/pipeline_data_tableau.pdf