



**Ministère de l'Enseignement Supérieur
Et de la Recherche Scientifique**

**Université de Tunis
Institut Supérieur de Gestion**

RAPPORT DE STAGE

SPECIALITE : BUSINESS INTELLIGENCE

**CONCEPTION ET DEVELOPPEMENT D'UNE SOLUTION BI POUR LA
DETECTION DES ANOMALIES DES DONNEES CLIENTELE DE LA
BANQUE**

Elaboré par :

Walid EL AOUN

Entreprise d'accueil :

Attijari Bank Tunisie

Encadrant pédagogique :

Mr. Lamjed Ben Said

Encadrant professionnel :

Mr. Nabil Zidi

Année universitaire

2022 – 2023

Dédicaces

Je dédie ce travail

A mes parents

A mon cher père *Sabri*, qui a su croire en moi dans les moments les plus délicats et qui m'a donné sa confiance totale, pour tous ses sacrifices pour que je puisse arriver à ce que je suis aujourd'hui.

A ma chère mère *Najla*, qui m'a toujours soutenu et encouragé dans les moments difficiles, pour les sacrifices qu'elle a faits pour moi, qui m'a permis d'évoluer dans les conditions les plus sereines face aux défis du quotidien. Je le suis redouble d'une éducation dont je suis fier.

A mes grands-parents

Une spéciale dédicace à *mes grands-parents*, qui sont vraiment présents dans le meilleur et dans le pire des situations à me donner leurs conseils précieux.

Une pensée très spéciale à mon grand-père *Sadok*, qu'il repose en paix.

A mon frère et ma sœur

Je dédie ce projet à mon frère *Sabri* et ma sœur *Hiba*, qui sont les deux personnes les plus chères à mes yeux qui ont toujours cru en moi, et auxquels je souhaite plein de réussite.

A tous mes proches et amis

Mes dédicaces sont offertes également à tous mes amis avec qui j'ai grandi et j'ai partagé et je partage les meilleurs moments.

Je dédie ce mémoire à tous ceux et celles qui comptent énormément pour moi et que je n'ai pas cité.

Remerciements

Je tiens dans un premier temps à remercier très sincèrement l'ensemble ***des employés et cadres de Attijari Bank*** pour leur accueil chaleureux et pour la jovialité de ce stage pour l'enrichissement aussi bien technique qu'humain. Je remercie Madame ***Mariem Ben Kacem*** qui m'a assisté durant toute la période de stage.

Je tiens à remercier Monsieur ***Nabil Zidi***, directeur à Attijari Bank de m'avoir accueilli durant ces 4 mois et de m'avoir donné le champ libre pour partager mes réflexions et mes propositions à chaque étape de cette aventure. J'espère avoir été digne de la confiance qu'il m'a accordée.

Je souhaite remercier Monsieur ***Lamjed Ben Said***, mon très respectable encadrant académique à ISG (Institut Supérieur de Gestion), pour ses orientations et ses remarques ciblées et pertinentes qui m'ont été très utiles pendant mon projet.

Je suis redevable à tous ***mes enseignants*** d'ISG pour les efforts fournis tout au long de ma formation au sein de l'institut et qui ont rendu ces 3 années exemplaires.

Mes humbles remerciements sont adressés aux ***membres de jury*** pour l'honneur qu'ils m'ont fait en acceptant de juger ce travail.

Table des matières

<i>Dédicaces</i>	<i>ii</i>
<i>Remerciements</i>	<i>iii</i>
<i>Table des matières.....</i>	<i>I</i>
<i>Liste des figures</i>	<i>IV</i>
<i>Liste des tableaux</i>	<i>XI</i>
<i>Liste des abréviations.....</i>	<i>XII</i>
<i>Introduction générale</i>	<i>1</i>
<i>Chapitre I. Cadre général du projet.....</i>	<i>3</i>
<i>Introduction</i>	<i>3</i>
<i>1. Organisme d'accueil.....</i>	<i>3</i>
1.1 Présentation Attijari Bank.....	3
1.2 Activités de l'entreprise	4
1.3 Organigramme Attijari Bank.....	4
<i>2. Cadre général du projet.....</i>	<i>5</i>
2.1 Contexte général du projet.....	5
2.2 Problématique	5
2.3 Les objectifs de Attijari Bank par rapport à ce projet	5
<i>3. Méthodologies de gestion des projets</i>	<i>6</i>
3.1 Méthodologie GIMSI [1]	6
3.2 Méthodologie SCRUM [2]	7
3.3 Etude Comparative entre GIMSI et SCRUM.....	8
<i>4. Choix des outils utilisés</i>	<i>9</i>
4.1 Outils ETL [3]	9
4.2 Outils de visualisation des données.....	10
<i>5. Spécification des besoins.....</i>	<i>13</i>
5.1 Besoins fonctionnels.....	13
5.2 Besoins non fonctionnels.....	13

6. Etude de l'existant	13
7. Planification du projet.....	14
Conclusion.....	14
<i>Chapitre II. Conception de la solution</i>	<i>15</i>
Introduction	15
1. Notions théoriques de la science des données	15
1.1 Définition de Business Intelligence.....	15
1.2 La chaîne décisionnelle.....	15
1.3 Les modèles multidimensionnels d'un Data Warehouse	16
2. Choix des indicateurs	18
2.1 Critères de choix	19
2.2 Sélection des indicateurs	19
3. Modélisation de l'entrepôt de données	19
4. Système du tableau de bord.....	20
Conclusion.....	22
<i>Chapitre III. Alimentation de l'entrepôt de données.....</i>	<i>23</i>
Introduction	23
1. Génération de données simulées.....	23
2. Réalisation du processus ETL	26
2.1. Création des Jobs, importation des fichiers sources et configuration de la connexion à la base de données ORACLE	26
2.2. Chargement des fichiers sources dans la base de données ORACLE	28
2.3. Application des règles de gestion pour transformer les données	33
2.3.1 Application des règles de gestion pour transformer les données liées aux adresses des clients.....	34
2.3.2 Application des règles de gestion pour transformer les données liées aux dates de naissance des clients.....	41
2.3.3 Application des règles de gestion pour transformer les données liées aux numéros de CIN48	
2.3.4 Application des règles de gestion pour transformer les données liées aux noms et prénoms des clients	53
2.3.5 Application des règles de gestion pour transformer les données liées aux numéros de téléphone des clients	58

2.3.6 Application des règles de gestion pour transformer les données liées aux adresses mails des clients.....	63
2.4. Charger les dimensions et les tables de faits	67
3.Conclusion.....	71
<i>Chapitre IV. Restitution des données.....</i>	72
Introduction	72
1. Connexion à la base de données, calcul des mesures de la table de faits et vérification de la structure du modèle de données.....	72
2. Page d'accueil du rapport	77
3. Analyse et interprétation des graphiques de la détection des anomalies des données clientèles de la banque	77
Conclusion.....	93
<i>Conclusion générale.....</i>	94
<i>Webographie.....</i>	96
Résumé	97
Abstract.....	97

Liste des figures

Figure 1 : Logo de Attijari Bank.....	3
Figure 2 : Organigramme de Attijari Bank	4
Figure 3 : Abréviation de l'acronyme GIMSI	6
Figure 4 : Les étapes de la méthode GIMSI	6
Figure 5 : Les étapes de la méthode SCRUM.....	7
Figure 6 : Diagramme de Gantt de notre projet	14
Figure 7 : Table de fait.....	16
Figure 8 : Table de dimension	17
Figure 9 : Modèle en étoile	17
Figure 10 : Modèle en flocon.....	18
Figure 11 : Modèle en constellation	18
Figure 12 : Data Warehouse	20
Figure 13 : Maquette tableaux de bords de la première page	21
Figure 14 : Maquette tableaux de bords de la deuxième page.....	21
Figure 15 : Maquette tableaux de bords de la troisième page	21
Figure 16 : Simulation du fichier Nom_Prenom.xls.....	24
Figure 17 : Simulation du fichier Email.xls.....	24
Figure 18 : Simulation du fichier Tel.xls.....	25
Figure 19 : Simulation du fichier CIN.xls	25
Figure 20 : Simulation du fichier Segmentation.xls	25
Figure 21 : Simulation du fichier Gestionnaires.xls	26
Figure 22 : Liste des jobs de l'ODS.....	27
Figure 23 : Métadonnées des fichiers Excel	27
Figure 24 : Connexion à la base de données ORACLE.....	27
Figure 25 : Espace de travail du job « adresse ».....	28
Figure 26 : Chargement de la table « adresse »	28
Figure 27 : Espace de travail du job « âge »	28
Figure 28 : Chargement de la table « âge »	29
Figure 29 : Espace de travail du job « CIN »	29
Figure 30 : Chargement de la table « CIN »	29
Figure 31 : Espace de travail du job « Email »	30

Figure 32 : Chargement de la table « EMAIL »	30
Figure 33 : Espace de travail du job « Noms et Prénoms »	30
Figure 34 : Chargement de la table « Nom_Prenom »	31
Figure 35 : Espace de travail du job « Tel »	31
Figure 36 : Chargement de la table « Tel »	31
Figure 37 : Espace de travail du job « CLIENT »	32
Figure 38 : Chargement de la table « Client »	32
Figure 39 : Ensemble des contextes des métadonnées et de la base de données ORACLE « Detect_Anomalies »	33
Figure 40 : Environnement de travail du job « T_Adresse »	34
Figure 41 : Table « Adresse »	34
Figure 42 : Structure de la table « Adresse »	35
Figure 43 : Fichier « Code_Postal.xls »	35
Figure 44 : Structure de « Code_Postal.xls »	35
Figure 45 : Composant « tMap_Adresse »	36
Figure 46 : Jointure des tables d'entrées	37
Figure 47 : Expression pour la règle de gestion qui consiste à avoir au moins un mot signifiant une adresse	37
Figure 48 : Expression pour la règle de gestion qui consiste à avoir au moins 3 mots dans les adresses	37
Figure 49 : Expression pour la règle de gestion qui consiste à avoir au moins un chiffre dans les adresses	38
Figure 50 : Expression pour la règle de gestion qui consiste à avoir au moins deux lettres dans chaque mot	38
Figure 51 : Traitement de la fonction « countLetters »	38
Figure 52 : Table « Adresse_correctes »	39
Figure 53 : Table « Anomalies_adresse »	39
Figure 54 : Composant tOracleOutput pour la table « Adresse_Correctes »	40
Figure 55 : Composant tOracleOutput pour la table « Anomalies_Adresse »	40
Figure 56 : Exécution du job « T_adresse »	41
Figure 57 : Chargement de la table « Anomalies_Adresses »	41
Figure 58 : Chargement de la table « Adresses_correctes »	41
Figure 59 : Environnement de travail du job « T_Age »	42

Figure 60 : Table « Age »	42
Figure 61 : Structure de la table « Age »	42
Figure 62 : Composant « tMap1 »	43
Figure 63 : Expression pour la règle de gestion qui consiste à avoir les dates de naissance sous la forme dd/mm/yyyy	43
Figure 64 : Composant « tFiltreRow »	43
Figure 65 : Composant « tMap2 »	44
Figure 66 : Expression pour la règle de gestion qui consiste à avoir les dates de naissance dont l'âge est compris entre 0 et 100.....	44
Figure 67 : Composant « tConvertType »	45
Figure 68 : Composant tOracleOutput pour la table « Age_Correctes ».....	45
Figure 69 : Composant « tMAp3 ».....	46
Figure 70 : Composant tOracleOutput pour la table « Anomalies_Age ».....	46
Figure 71 : Exécution du job « T_age »	47
Figure 72 : Chargement de la table « Age_correctes ».....	47
Figure 73 : Chargement de la table « Anomalies_Age ».....	48
Figure 74 : Environnement de travail du job « T_CIN ».....	48
Figure 75 : Table « CIN ».....	49
Figure 76 : Structure de la table « CIN ».....	49
Figure 77 : Composant tMap pour la transformation des numéros de CIN.....	50
Figure 78 : Expression pour la règle de gestion qui consiste à avoir les numéros de CIN qui commence que par 0 et 1 et leur longueur est de 8 chiffres	50
Figure 79 : Expression pour la règle de gestion qui consiste à avoir les numéros de CIN qui contiennent que des numéros	51
Figure 80 : Composant tOracleOutput pour la table « CIN_Correct ».....	51
Figure 81 : Composant tOracleOutput pour la table « Anomalies_CIN »	52
Figure 82 : Exécution du job « T_CIN »	52
Figure 83 : Chargement de la table « CIN_correctes »	53
Figure 84 : Chargement de la table « Anomalies_CIN ».....	53
Figure 85 : Environnement de travail du job « T_Nom_Prenom »	54
Figure 86 : Table « Nom_Prenom »	54
Figure 87 : Structure de la table « Nom_Prenom »	54
Figure 88 : Composant « tMap_Nom_Prenom ».....	55

Figure 89 : Expression pour la règle de gestion qui consiste à avoir les noms et prénoms non null.....	55
Figure 90 : Composant tOracleOutput pour la table « Nom_Prenom_Corrects ».....	56
Figure 91 : Composant tOracleOutput pour la table « Anomalies_Nom_Prenom ».....	56
Figure 92 : Exécution du job « T_Nom_Prenom ».....	57
Figure 93 : Chargement de la table « Nom_Prenom_corrects ».....	57
Figure 94 : Chargement de la table « Anomalies_Nom_Prenom ».....	57
Figure 95 : Environnement de travail du job « T_Tel »	58
Figure 96 : Table « Tel »	59
Figure 97 : Structure de la table « Tel »	59
Figure 98 : Composant « tMap_Tel ».....	59
Figure 99 : Fonction java pour standardiser les numéros de téléphone.....	60
Figure 100 : Application de la fonction « cleanPhoneNumber » pour les numéros de téléphone fiable	60
Figure 101 : Application de la fonction « cleanPhoneNumber » pour les numéros de téléphone portable	60
Figure 102 : Expression pour la règle de gestion qui consiste à avoir les numéros de téléphone dont la longueur est exactement 8 chiffres et qui ne commence pas par 0 ou 3ou 4ou bien 7	61
Figure 103 : Composant tOracleOutput pour la table « Tel_Corrects ».....	61
Figure 104 : Composant tOracleOutput pour la table « Anomalies_Tel »	61
Figure 105 : Exécution du job « T_Tel ».....	62
Figure 106 : Chargement de la table « Tel_correctes »	62
Figure 107 : Chargement de la table « Anomalies_Tel »	62
Figure 108 : Environnement de travail du job « T_Email »	63
Figure 109 : Table « Email »	63
Figure 110 : Structure de la table « Email »	64
Figure 111 : Composant « tMap_Mail ».....	64
Figure 112 : Expression pour la règle de gestion qui consiste à avoir les adresses mail qui ne contiennent pas des caractères spéciaux	65
Figure 113 : Expression pour la règle de gestion qui consiste à avoir les adresses mail qui contienne un nom de service mail	65

Figure 114 : Expression pour la règle de gestion qui consiste à avoir les adresses mail qui commence avec un caractère alphabétique	65
Figure 115 : Composant tOracleOutput pour la table « Anomalies_Email ».....	65
Figure 116 : Composant tOracleOutput pour la table « Email_Corrects ».....	66
Figure 117 : Exécution du job « T_Email »	66
Figure 118 : Chargement de la table « Email_corrects ».....	66
Figure 119 : Chargement de la table « Anomalies_Email »	67
Figure 120 : Chargement de la dimension « Dim_Gestionnaire » à partir du fichier source	67
Figure 121 : Chargement de la dimension « Dim_Agence » à partir du fichier source ...	67
Figure 122 : Chargement de la dimension « Dim_Segmentation_Client » à partir du fichier source	68
Figure 123 : Chargement de la dimension « Dim_Client_Anomalies ».....	68
Figure 124 : Composant « tMap_Dim_Client_anomalies »	69
Figure 125 : Expression pour récupérer les données incorrectes et transformer les données correctes en « null »	69
Figure 126 : Chargement de la table des faits « Fact_Anomalies »	69
Figure 127 : Job master pour l'ODS	70
Figure 128 : Job master pour les jobs des règles de gestion	70
Figure 129 : Job master pour les jobs des dimensions et de la table de fait	70
Figure 130 : Clés étrangères de la table des faits.....	71
Figure 131 : Importation des données du Data Warehouse	73
Figure 132: Configuration de la connexion à la base de données Oracle.....	73
Figure 133 : Modélisation en étoile du Data Warehouse dans Power BI	76
Figure 134 : Page d'accueil du rapport.....	77
Figure 135 : Page « Information générales »	78
Figure 136 : Répartition des fiches clients selon la direction régional, le groupe et le nom des agences	78
Figure 137 : Répartition des fiches clients selon le marché, le segment NMR et le sous segment.....	79
Figure 138 : Répartition des fiches clients selon leurs professions	79
Figure 139 : Répartition des fiches clients selon leurs spécificités	80
Figure 140 : Répartition des fiches clients selon leur pays, région et ville	80

Figure 141 : Page « Anomalies ».....	81
Figure 142 : Répartition des fiches clients ayant des anomalies selon la direction régional, le groupe et le nom des agences	82
Figure 143 : Nombre de fiches clients ayants des anomalies par gestionnaires	83
Figure 144 : Fiches Clients	83
Figure 145 : Bookmark « Adresse »	84
Figure 146 : Répartition des fiches clients ayant des anomalies dans leurs adresses selon le segment NMR	84
Figure 147 : Répartition des fiches clients ayant des anomalies dans leurs adresses selon la direction, le groupe et le nom d'agence	85
Figure 148 : Répartition des fiches clients ayant des anomalies dans leurs adresses selon le marché, la profession et le sous segment	85
Figure 149 : Répartition des fiches clients ayant des anomalies dans leurs adresses selon leurs spécificités	86
Figure 150 : Filtre des adresses	86
Figure 151 : Bookmark « Date de naissance »	87
Figure 152 : Bookmark « Email »	87
Figure 153 : Bookmark « CIN »	88
Figure 154 : Bookmark « Téléphone »	88
Figure 155 : Bookmark « Nom et prénom ».....	89
Figure 156 : Page « Evolution du nombre des anomalies au cours du temps »	89
Figure 157 : Nombre de fiche clients par date de modification.....	90
Figure 158 : Bookmark « Evolution du nombre des anomalies adresse au cours du temps »	90
Figure 159 : Bookmark « Evolution du nombre des anomalies de date de naissance au cours du temps »	91
Figure 160 : Bookmark « Evolution du nombre des anomalies dans les numéros de téléphones au cours du temps ».....	91
Figure 161 : Bookmark « Evolution du nombre des anomalies dans les numéros de CIN au cours du temps ».....	91
Figure 162 : Bookmark « Evolution du nombre des anomalies dans les nom et prénoms au cours du temps ».....	92

Figure 163 : Bookmark « Evolution du nombre des anomalies email au cours du temps »	92
Figure 164 :Filtres agence et date	93

Liste des tableaux

Tableau 1 : Comparatif des méthodes GIMSI et SCRUM	8
Tableau 2 : Comparatif des outils d'ETL « Talend Open Studio for Data Integration » et « SQL Server Integration Services ».....	9
Tableau 3 : Comparatif des outils « Power BI Desktop » , « Tableau Desktop » et « Qlik Sense ».....	11
Tableau 4 :Formules DAX des mesures dans Power BI.....	74

Liste des abréviations

KPI : abréviation anglaise Key Performance Indicator

ETL : abréviation anglaise de Extract-transform-load

PFE : Projet Fin d'Etudes

ODS : abréviation anglaise de Operational Data Store

CIN : carte d'identité nationale

ISG : Institut Supérieur de Gestion

Introduction générale

La gouvernance des données et la qualité des données sont des éléments clés pour toute organisation qui souhaite prendre des décisions efficaces et basées sur des données fiables. Cependant, dans le contexte de la banque, il est souvent difficile de garantir que les données clientèles sont complètes, précises et à jour, ce qui peut entraîner des conséquences néfastes sur les décisions prises par les responsables de la banque.

Dans ce contexte, ce PFE vise à améliorer la qualité des données clientèles de la banque en mettant en place une solution BI pour détecter les anomalies dans ces données. Cette solution permettra d'identifier les erreurs et les incohérences et de les signaler aux responsables de la banque. Cette approche proactive permettra d'assurer que les données utilisées pour prendre des décisions sont fiables et précises.

Objectif

Ce projet d'étude a pour but d'exploiter les données issues du système d'information d'Attijari Bank, afin d'établir un rapport sur les anomalies des données clientèles de la banque. Ce rapport permettra de fournir aux décideurs du département Référentiel Client (RC), des indicateurs pertinents pour les aider à piloter leurs activités et mettre en place une veille permanente.

Solution proposée

Afin de répondre aux besoins fonctionnels exprimés par la banque dans le cadre de l'activité de détection des anomalies, nous avons mis en place une solution BI (Business intelligence ou informatique décisionnelle en français), qui se chargera d'extraire, traiter et restituer les données dans un tableau de bord interactif en adéquation avec les besoins métiers du département RC.

Organisation du rapport

Le présent rapport se compose de quatre chapitres ainsi qu'une conclusion générale :

-Le premier chapitre du rapport intitulé « Cadre général du projet » est consacré à la présentation du cadre du projet, l'entreprise d'accueil, les méthodologies et les outils utilisés, la spécification des besoins auxquels doit répondre ce projet, l'étude de l'existant et la planification du projet.

-Le deuxième chapitre du rapport intitulé « Conception de la solution», présente un rappel des notions théoriques utilisées dans le cadre de ce projet ainsi que le choix des indicateurs , la conception du Data Warehouse et le prototype de nos tableaux de bord.

-Le troisième chapitre du rapport intitulé « Alimentation de l'entrepôt de données », retranscrit les différentes étapes réalisées tout au long du processus ETL.

-Le quatrième chapitre du rapport intitulé « Restitution des données » est destiné à l'analyse et l'interprétation graphique des données clientèles de la banque sous forme de Storytelling.

-Enfin, nous achèverons le rapport par une conclusion générale retracant les différentes étapes réalisées du projet.

Chapitre I. Cadre général du projet

Introduction

Le domaine banquier représente l'un des piliers les plus importants dans l'économie tunisienne. En premier lieu, ce chapitre explicitera davantage ce secteur en mettant en lumière Attijari Bank, qui compte parmi les premières banques en Tunisie, en citant son historique, ses activités et son organigramme. En deuxième lieu, nous présenterons une comparaison entre les méthodologies et les outils de travail. Finalement, nous développerons la problématique de ce projet de fin d'études en se focalisant sur l'objectif de Attijari Bank tiré de cette concrétisation.

1. Organisme d'accueil

1.1 Présentation Attijari Bank

Attijari Bank Tunisie est née en 2005 suite à la privatisation de la banque nationale «Banque du Sud ». Le capital de l'Etat Tunisien dans les actions de la Banque du Sud a été acquis par le groupe Andalou maghrébin (conglomérat maghrébin espagnol formé par le groupe Attijari Wafa Bank et Santander Espagne).

La banque dispose aujourd'hui d'un réseau bancaire de 198 agences et centres d'affaires répartis sur tout le territoire tunisien. Chez Attijari Bank, les institutions financières sont présentes dans tous les domaines d'activité. Elle opère à travers des filiales sur les marchés des particuliers, des professionnels et des entreprises. En plus de l'intermédiation, ils se concentrent sur la location et la gestion d'actifs. Il s'agit des domaines du conseil, de l'assurance, de la bourse, du capital-risque et de la philanthropie de risque.



Figure 1 : Logo de Attijari Bank

1.2 Activités de l'entreprise

Les principales activités exercées par Attijari Bank en Tunisie sont :

- Une banque de détail destinée aux particuliers et les petites et moyennes entreprises,
- Une banque de Financement de l'économie et d'Investissement destinée pour la grande entreprise et les institutionnels
- A travers ses filiales spécialisées, la banque assure les activités d'intermédiation, de leasing et de bancassurance.

1.3 Organigramme Attijari Bank

La figure 2 ci-dessous représente l'organigramme de Attijari Bank, structuré selon 3 niveaux :

- La direction générale
- Les pôles et les départements
- Les directions et les sous-directions

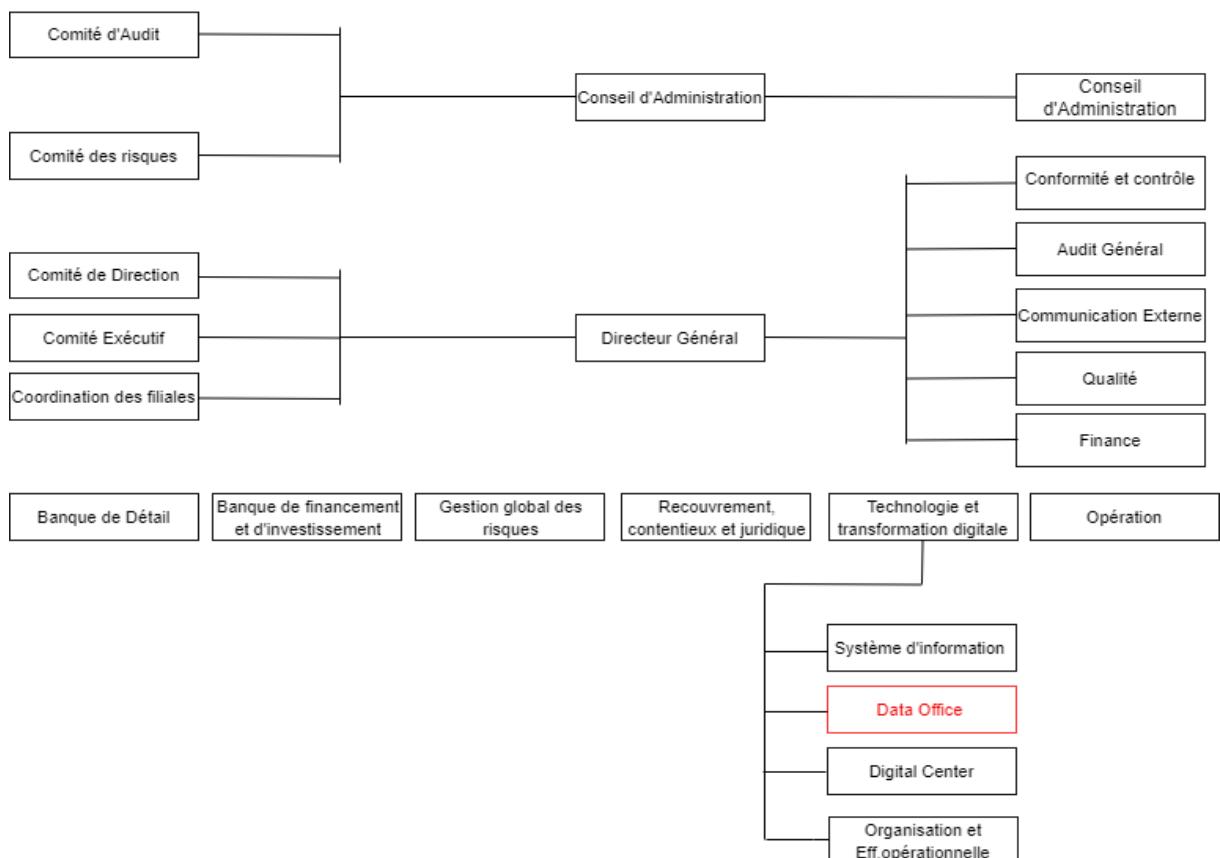


Figure 2 : Organigramme de Attijari Bank

Ce stage a été effectué au sein du département Data Office. Ce département est constitué d'une équipe de consultants décisionnels et de Data Scientists, qui ont pour mission d'assurer la gouvernance des données relatives aux clients de la banque, leurs niveaux de qualités et de leurs utilisations analytiques dans l'objectif d'offrir aux clients de la banque des services plus personnalisés et plus pertinents grâce à la modélisation analytique pour allier le meilleur de l'humain et du digital.

2. Cadre général du projet

2.1 Contexte général du projet

Ce projet d'étude a pour but d'exploiter les données issues du système d'information d'Attijari Bank, afin d'établir un rapport sur les anomalies des données clientèles de la banque. Ce rapport permettra de fournir aux décideurs du département Référentiel Client (RC), des indicateurs pertinents pour les aider à piloter leurs activités et mettre en place une veille permanente.

2.2 Problématique

Manipuler les données clientèle de la banque en appliquant le processus ETL (abréviation anglaise de Extract-Transform-Load) qui consiste à l'extraction de données provenant de différents fichiers Excel, faire les transformations nécessaires de ces données et les charger dans la base de données à travers l'outils d'intégration Talend, afin de les exploiter dans la mise en place de mesures et KPIs (abréviation anglaise Key Performance Indicator) qui aident à détecter les anomalies et améliorer la qualité des données.

2.3 Les objectifs de Attijari Bank par rapport à ce projet

Plusieurs points positifs peuvent être tirés de la réalisation d'un tel projet parmi lesquels l'amélioration de la qualité des données et également la détection des anomalies dans les données clientèle de la banque.

3. Méthodologies de gestion des projets

Déterminer la bonne méthode de travail pour la prise de décision dans le cadre des projets professionnels est une étape importante et une stratégie extrêmement efficace pour organiser le travail et définir les différentes phases et objectifs qui conduisent à une bonne gestion de projet. En effet, il est nécessaire d'avoir une bonne approche dès le début du projet, car plusieurs approches se sont révélées infructueuses lorsqu'il s'agit de répondre aux attentes besoins métiers. Par conséquent, il est nécessaire d'étudier toutes les méthodes existantes et de choisir celle qui est la plus adaptée à notre projet.

3.1 Méthodologie GIMSI [1]

GIMSI est une méthode coopérative de conception de système de pilotage à base de tableaux de bord, qui permet d'établir le rôle et la responsabilité de tous décideurs confrontés au risque et à la complexité d'un projet décisionnel.

La figure 3 ci-dessous montre l'abréviation de l'acronyme GIMSI :

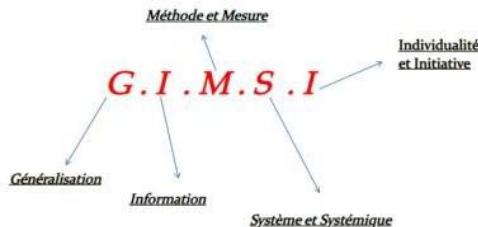


Figure 3 : Abréviation de l'acronyme GIMSI

La méthode GIMSI comporte 10 étapes bien déterminées et qui sont reparties en 4 phases thématiques :

Phase	N°	Etape	Objectifs
Identification <i>Quel est le contexte ?</i>	1	Environnement de l'entreprise	Analyse de l'environnement économique et de la stratégie de l'entreprise afin de définir le périmètre et la portée du projet
	2	Identification de l'entreprise	Analyse des structures de l'entreprise pour identifier les processus, activités et acteurs concernés
Conception <i>Que faut-il faire ?</i>	3	Définition des objectifs	Sélection des objectifs tactiques de chaque équipe
	4	Construction du tableau de bord	Définition du tableau de bord de chaque équipe
Mise en œuvre <i>Comment le faire ?</i>	5	Choix des indicateurs	Choix des indicateurs en fonction des objectifs choisis
	6	Collecte des informations	Identification des informations nécessaires à la construction des indicateurs
Amélioration permanente	7	Le système de tableau de bord	Construction du système de tableaux de bord, contrôle de la cohérence globale
	8	Le choix des progiciels	Élaboration de la grille de sélection pour le choix des progiciels adéquats
	9	Intégration et déploiement	Implantation des progiciels, déploiement à l'entreprise
	10	Audit	Suivi permanent du système <i>Le système correspond-il</i>

Figure 4 : Les étapes de la méthode GIMSI

3.2 Méthodologie SCRUM [2]

La méthodologie SCRUM est une méthode de gestion de projet de type Agile. Son objectif est d'améliorer la productivité des équipes, tout en permettant une optimisation du produit en développement grâce à des feedbacks réguliers avec les utilisateurs finaux. C'est une approche dynamique et participative de la conduite du projet, qui garantit pour le client le juste équilibre entre l'investissement prévu et le produit finalement livré. Cette méthode fait intervenir trois rôles complémentaires dans l'implication du projet :

- Le Scrum Master, dont le rôle est de promouvoir et de soutenir le projet tel que défini dans le Scrum Guide. Il le fait en aidant tout le monde à comprendre la théorie, les pratiques, les règles et les valeurs de la méthodologie Scrum.
- Le Product Owner, qui est chargé de faire le pont entre la partie métier et la partie technique du projet. Il est le relai entre le client et l'équipe de développement. C'est lui qui est responsable de la rédaction des user stories et de la maintenance du Product Backlog.
- L'équipe de développement, qui doit transformer les besoins exprimés en fonctionnalités utilisables.

La figure 5 ci-dessous schématise le déroulement d'un projet SCRUM :



Figure 5 : Les étapes de la méthode SCRUM

3.3 Etude Comparative entre GIMSI et SCRUM

Le tableau 1 permet de comparer les différentes phases des méthodes GIMSI et SCRUM :

Tableau 1 : Comparatif des méthodes GIMSI et SCRUM

	GIMSI	SCRUM
Cycle de vie	Non cyclique	Itératif
Phase 1	Identification : -Identifier l'entreprise et l'environnement de travail.	Définition des exigences : -Définir les objectifs finaux à atteindre.
Phase 2	Conception : -Définir les objectifs, collecter les données pour la conception du tableau de bord, choisir les indicateurs qui répondent aux besoins définis et créer le système de tableau de bord.	Backlog projet: -Définir les actions à réaliser pour atteindre l'objectif final.
Phase 3	Mise en œuvre : -Choisir les progiciels, faire l'intégration et déploiement du projet.	Planification des sprints : -Ranger dans le bon ordre les grands thèmes de travail.
Phase 4	Amélioration permanente : -Audit périodique du système de tableau de bord.	Sprints : -Les phases de production intenses qui permettent d'aboutir à un livrable intermédiaire

En s'appuyant sur cette étude comparative, nous avons opté pour la méthode GIMSI car c'est la méthodologie qui s'adapte le mieux à la réalisation de notre projet décisionnel et à l'environnement de la banque. Cette méthode va permettre aux décideurs et aux gestionnaires de prendre des décisions stratégiques et de réaliser le projet sans embûche.

4. Choix des outils utilisés

4.1 Outils ETL [3]

Nous avons réalisé un comparatif des outils d'ETL étudiés durant notre formation universitaire pour savoir lequel est le plus adapté aux critères d'évaluation d'un projet BI. Le tableau 2 permet de comparer les outils d'ETL « Talend Open Studio for Data Integration » et « SQL Server Integration Services » :

Tableau 2 : Comparatif des outils d'ETL « Talend Open Studio for Data Integration » et « SQL Server Integration Services »

		
Systèmes d'exploitation	Windows Linux MacOs	Windows
Développeur	Talend	Microsoft
Licence	Open Source	Commerciale
Avantages	-Propose plus de 600 composants et connecteurs pour se connecter facilement avec les principaux systèmes de gestion de base de données	-Permet de gérer les synchronisations, les migrations de données, leur transformation et leur intégration dans le Data Warehouse. -Possibilité de créer des connecteurs sur-mesure

	<ul style="list-style-type: none"> -Offre une représentation visuelle des flux de données pour l'organisation des jobs -Permet de planifier l'exécution des jobs et de gérer les erreurs 	<ul style="list-style-type: none"> -Possibilité d'exécuter de nombreux processus en parallèle
Inconvénients	<ul style="list-style-type: none"> -Nécessité de connaître le langage Java si l'on souhaite utiliser des fonctions avancées -Difficulté d'utilisation pour la recherche d'un composant spécifique -Logiciel assez lourd 	<ul style="list-style-type: none"> -Impossibilité de dupliquer un flux existent. -SIS ne peut être installé uniquement sur un système d'exploitation Windows via l'environnement de développement Visual Studio -Problèmes de performances liés à utilisation élevée de la mémoire vive

En s'appuyant sur cette étude comparative, nous avons choisi de travailler avec Talend Open Studio for Data Integration du fait de sa simplicité d'utilisation et de la richesse de sa documentation.

4.2 Outils de visualisation des données

Comme pour le choix de l'outil ETL, nous avons réalisé un comparatif des outils utilisés pour la visualisation de données afin de savoir lequel est le plus adapté à nos besoins. Le tableau 3 de comparer les outils « Power BI Desktop » , « Tableau Desktop » et « Qlik Sense » [4] :

Tableau 3 : Comparatif des outils « Power BI Desktop », « Tableau Desktop » et « Qlik Sense »

	 Power BI	 + a b e a u	
Développeur	Microsoft	Tableau Software	Qlik
Système d'exploitation	Windows	Windows MacOs	Windows
Licence	Propriétaire	Propriétaire	Propriétaire
Avantages	<ul style="list-style-type: none"> -Interface intuitive. -Compatible avec de nombreuses sources de données -Possibilité de personnaliser son tableau de bord en important des thèmes ou des visuels créés par la communauté. -Power BI est composé de Power Query qui possède des fonctionnalités avancées de transformation des données 	<ul style="list-style-type: none"> -Possibilité de créer rapidement des graphiques à l'aide d'un simple glisser-déposer. -Compatible avec de nombreuses sources de données -Possède des fonctionnalités avancées de storytelling -Utilise le langage MDX qui possède une syntaxe appropriée à l'interrogation et la manipulation des données multidimensionnelles mémorisées dans un cube OLAP. 	<ul style="list-style-type: none"> -Permet de construire des tableaux de bord personnalisés. -Permet l'analyse, l'interprétation et le partage de données complexes. -À la capacité de s'intégrer à un système d'applications d'entreprise avec de nombreuses API.

	-Utilise le langage DAX qui permet de créer des mesures, des colonnes et des tables à l'aide de formules semblables à celles utilisées dans Excel.		
Inconvénients	<ul style="list-style-type: none"> -Les tableaux de bord et rapports ne peuvent être partagés qu'avec des personnes ayant des emails sur le même domaine. -Ne peut pas mélanger les données issues de multiples sources de données. 	<ul style="list-style-type: none"> -Le coût de la licence est élevé et la version gratuite - Tableau Public est limitée en termes de fonctionnalités. -Ne propose pas de capacités d'IA ou de Machine Learning. 	<ul style="list-style-type: none"> -Interface peu intuitive. -L'analyse des données en temps réel est parfois impossible avec QlikView.

En s'appuyant sur cette étude comparative, nous avons choisi de travailler avec Power BI du fait de sa simplicité d'utilisation et des fonctionnalités offertes par le langage DAX et l'éditeur de données Power Query.

5. Spécification des besoins

Il est primordial de fixer les besoins fonctionnels et non fonctionnels afin de satisfaire les besoins métiers de notre projet.

5.1 Besoins fonctionnels

Les fonctionnalités nécessaires pour notre projet sont les suivantes :

-Élaboration d'un modèle conceptuel de l'entrepôt de données.

-Collecte et génération de données fictives en relation avec le modèle multidimensionnel établi.

-Mise en œuvre du processus ETL pour extraire, transformer et charger les données utiles dans l'entrepôt de données.

-Création d'un rapport comprenant des tableaux de bord permettant l'analyse et l'interprétation graphique des données clients de la banque, en particulier pour détecter les anomalies.

5.2 Besoins non fonctionnels

Nous devons également prendre en compte les exigences non fonctionnelles suivantes pour notre projet :

-Le Dashboard doit inclure une page d'accueil qui expose clairement l'objectif et la structure du rapport.

-Les pages de l'analyse doivent être bien organisées et faciles à comprendre, ce qui nécessite la mise en place d'un menu de navigation.

-Le rapport doit respecter la charte graphique d'Attijari Bank en utilisant les couleurs correspondant à ses codes.

6. Etude de l'existant

Afin d'avoir une bonne qualité des données, la banque a commencé à appliquer plusieurs règles de gestion à ses données clients à l'aide du logiciel « Talend Open Studio for Data Integration ». Pour mieux appréhender les anomalies des données clients, il est indispensable de mettre en place un tableau de bord pour pouvoir identifier la répartition de ces anomalies au sein du réseau de la banque.

C'est pourquoi nous allons modéliser un DataMart et reprendre l'application des règles de gestion afin d'obtenir des anomalies dans les données clients puis nous les afficherons dans un tableau de bord à l'aide de l'outil Microsoft "Power BI".

7. Planification du projet

Après avoir défini les différents besoins du projet, il faut répartir les différentes tâches à réaliser et visualiser l'avancement du projet grâce au diagramme de Gantt. La figure 6 montre un diagramme de Gantt qui modélise la répartition des tâches durant ce projet :

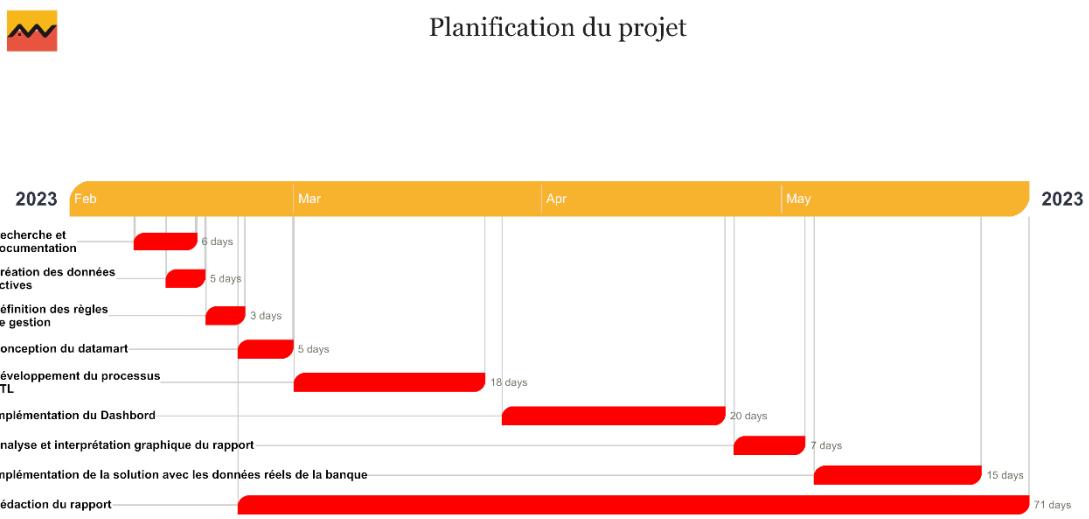


Figure 6 : Diagramme de Gantt de notre projet

Conclusion

Dans ce premier chapitre, nous avons parlé du cadre général du projet en présentant le contexte du projet, la problématique et les objectifs de Attijari Bank par rapport à ce projet. Ensuite nous avons choisi la méthodologie et les outils de travail. Finalement nous avons présenté les besoins (fonctionnels et non fonctionnels) de la banque et mis une planification du projet à l'aide d'un diagramme de Gantt.

Chapitre II. Conception de la solution

Introduction

Le deuxième chapitre de ce document abordera les concepts théoriques relatifs à la Business Intelligence pour permettre une meilleure compréhension des outils et des méthodologies utilisés dans le cadre de projets d'analyse de données. Ensuite nous allons présenter la conception de notre solution en commençant par le choix des indicateurs, la conception du Data Warehouse et la présentation du prototype du tableau de bord.

1. Notions théoriques de la science des données

1.1 Définition de Business Intelligence

La Business Intelligence (BI) regroupe un ensemble de méthodes et d'outils destinés à collecter, consolider, analyser et présenter les données de l'entreprise, afin de faciliter la prise de décision. Les technologies utilisées dans la BI incluent les entrepôts de données, les rapports, les tableaux de bord et les analyses prédictives, qui fournissent des informations stratégiques et opérationnelles aux décideurs de l'entreprise. L'objectif de la BI est d'améliorer la visibilité et la compréhension des données pour optimiser les performances et la gestion des activités de l'entreprise.

1.2 La chaîne décisionnelle

La chaîne décisionnelle est le processus de traitement de l'information qui permet de convertir les données collectées en informations exploitables pour la prise de décisions. Cette chaîne est généralement composée de quatre étapes consécutives :

-La première étape consiste à extraire et à collecter les données des systèmes de production de l'entreprise, puis à les transformer et à les charger dans une base de données à l'aide du processus ETL (Extract, Transform, Load). Les données nettoyées sont stockées dans

un entrepôt de données géré par un SGBDR (Système de Gestion de Bases de Données Relationnelles) tel que MySQL, Microsoft SQL Server, Oracle Database ou PostgreSQL.

-La deuxième étape consiste à stocker les données transformées dans l'entrepôt de données pour une utilisation ultérieure à des fins décisionnelles.

-La troisième étape est consacrée à la restitution des données à l'aide d'outils de reporting tels que Power BI, Tableau ou QlikView.

-La quatrième et dernière étape de la chaîne décisionnelle consiste à exploiter les données. Les tableaux de bord créés précédemment peuvent maintenant être utilisés par les utilisateurs finaux ou les experts en analyse.

1.3 Les modèles multidimensionnels d'un Data Warehouse

La modélisation multidimensionnelle est une méthode couramment utilisée pour concevoir un data Warehouse efficace. Cette méthode implique l'utilisation de deux types de tables : la table de faits et la table de dimension.

La table de faits (figure 7) correspond à la table centrale du modèle multidimensionnel, elle contient des informations quantitatives et mesurables sur un événement spécifique. Cela peut inclure des informations telles que des ventes, des revenus ou des quantités de produits vendus. Les attributs de la table de faits comprennent généralement des mesures, des dates et des clés étrangères qui se rapportent aux tables de dimension.

Fait
<u>Clé de dimension1</u>
<u>Clé de dimension2</u>
Mesure1
Mesure2
Mesure3.....

Figure 7 : Table de fait

D'autre part, les tables de dimension (figure 8) fournissent des informations contextuelles sur les données contenues dans la table de faits. Ces tables contiennent des informations qui

décrivent chaque ligne d'enregistrement de la table et qui peuvent être utilisées pour filtrer, regrouper ou agréger les données de la table de faits. Les attributs de la table de dimension comprennent souvent des informations telles que le temps, le lieu, le produit ou le client.

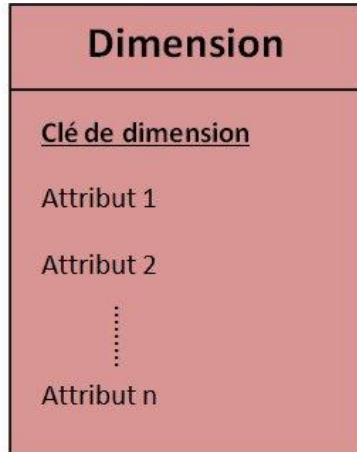


Figure 8 : Table de dimension

Le modèle utilisé pour concevoir un data Warehouse dépendra des besoins spécifiques de l'entreprise ou de l'organisation. Certains modèles couramment utilisés comprennent le modèle en étoile, le modèle en flocon de neige et le modèle constellation.

Le modèle en étoile (figure 9) est le plus simple et le plus couramment utilisé, avec une table de faits centrale reliée à plusieurs tables de dimension.

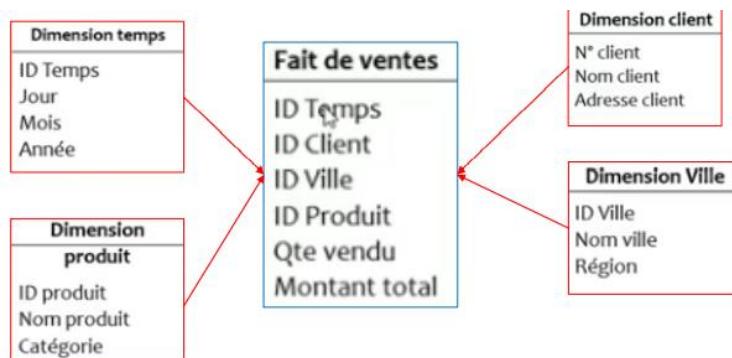


Figure 9 : Modèle en étoile

Le modèle en flocon de neige (figure 10) est similaire au modèle en étoile, mais les tables de dimension sont normalisées pour réduire la redondance des données.

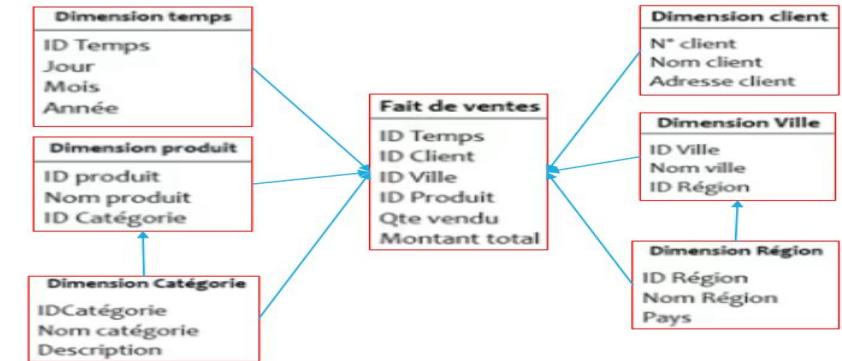


Figure 10 : Modèle en flocon

Le modèle constellation (figure 11) est le plus complexe, avec plusieurs tables de faits partageant des tables de dimension communes.

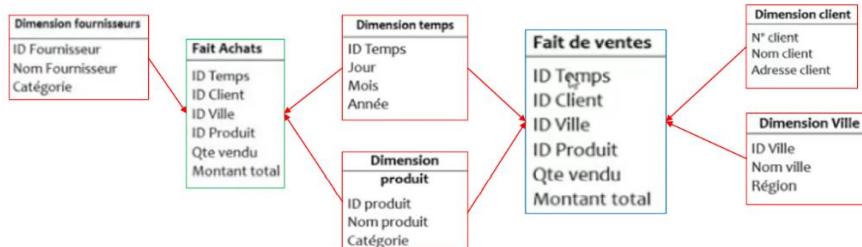


Figure 11 : Modèle en constellation

En résumé, la modélisation multidimensionnelle utilise des tables de faits et de dimension pour concevoir un data Warehouse efficace. Les attributs de la table de faits sont quantitatifs et mesurables, tandis que les attributs de la table de dimension fournissent un contexte pour les données de la table de faits. Les modèles utilisés pour concevoir un data Warehouse varient en complexité et en structure, selon les besoins de l'entreprise ou de l'organisation.

2. Choix des indicateurs

La phase de conception atteint un point crucial lors de cette étape, où les indicateurs clés de performance sont déterminés. Pour ce faire, nous avons consulté à plusieurs reprises les analystes de la banque afin d'établir des indicateurs pertinents.

2.1 Critères de choix

Un indicateur est un élément ou une variable qui peut être objectivement mesuré et qui est utilisé pour évaluer de manière fiable les changements réalisés ou les progrès accomplis par un projet dans l'atteinte de ses résultats et objectifs. Les indicateurs doivent présenter les caractéristiques suivantes :

-Fiabilité : ils doivent pouvoir être mesurés de manière cohérente dans le temps et de la même manière par différents observateurs.

-Précision : ils doivent pouvoir être mesurés de manière claire à l'aide de données quantitatives ou qualitatives.

2.2 Sélection des indicateurs

Dans cette étape, nous allons fixer les indicateurs qui seront utiliser dans les visuels du tableau de bord.

-Le taux des anomalies par rapport au nombre totales des fiches clients

-Le taux des anomalies par rapport au nombre total des anomalies

-Le nombre de fiches clients ayant des anomalies

-Le nombre de fiches client correctes

-Le segment avec le plus de fiches clients ayant des anomalies (pour chaque type d'anomalies)

-Le marché avec le plus de fiches clients ayant des anomalies (pour chaque type d'anomalies)

-La profession avec le plus de fiches clients ayant des anomalies (pour chaque type d'anomalies)

3. Modélisation de l'entrepôt de données

L'une des étapes les plus importantes dans notre projet décisionnel consiste à concevoir le Data Warehouse en choisissant le modèle de conception multidimensionnel le plus approprié à notre sujet d'analyse. Après avoir examiné les anomalies liées au données clientèles de la banque, nous avons décidé d'opter pour un modèle en étoile pour concevoir notre Data Warehouse. La gestion des anomalies de données clientèles de la banque peut être gérée

efficacement avec un modèle en étoile. La figure 12 présente le modèle en étoile que nous avons adopté pour notre Data Warehouse. Nous avons utilisé « diagrams.net » pour la modélisation.

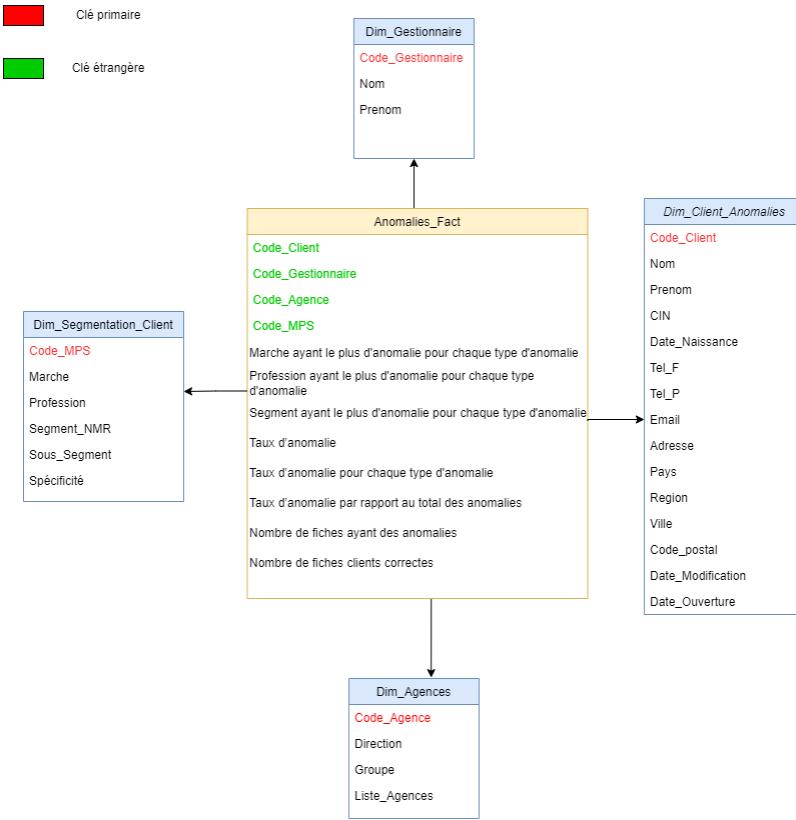


Figure 12 : Data Warehouse

4. Système du tableau de bord

Dans cette partie nous parlerons principalement du système de notre solution. En effet l'un des objectifs qu'on s'est fixé lors de cette étape est l'ergonomie. Un tableau de bord ergonomique devrait assurer un accès rapide à l'information.

C'est dans ce contexte qu'on a voulu utiliser la notion des bookmarks pour pouvoir naviguer entre les différents tableaux de bord dans une même page à l'aide de boutons.

-Le prototype des tableaux de bord :

Suite à l'étude approfondie des besoins de la banque nous avons élaboré quelques maquettes à suivre dans la création de nos tableaux de bord comme le montre les figures de 13 à 15.



Figure 13 : Maquette tableaux de bord de la première page



Figure 14 : Maquette tableaux de bord de la deuxième page

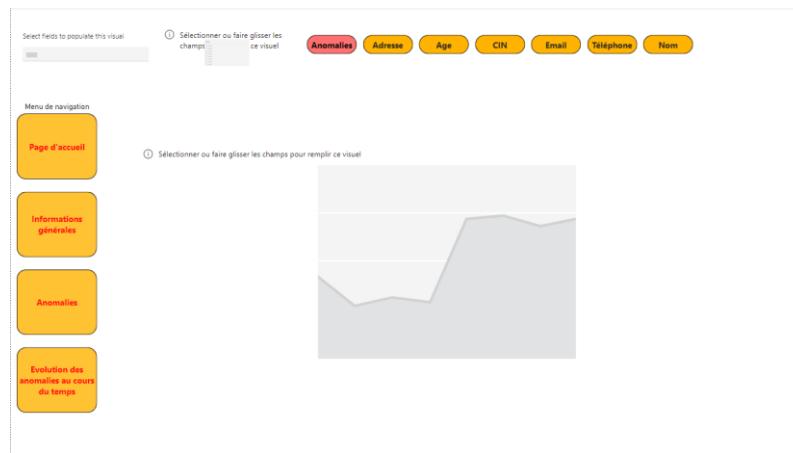


Figure 15 : Maquette tableaux de bord de la troisième page

Conclusion

Dans ce deuxième chapitre, nous avons acquis une compréhension théorique des fondements de la Business Intelligence pour pouvoir ensuite présenter la conception de notre solution en commençant par le choix des indicateurs, la conception du Data Warehouse et la présentation du prototype du tableau de bord.

Chapitre III. Alimentation de l'entrepôt de données

Introduction

Dans ce troisième chapitre, nous allons commencer la première étape d'un projet décisionnel, qui consiste à concevoir et à alimenter l'entrepôt de données à l'aide du processus ETL. Cette étape permettra de fournir des données agrégées, structurées, historiques et non volatiles au Data Warehouse, afin de pouvoir effectuer des analyses de détection des anomalies des données clientèles de la banque.

1. Génération de données simulées

Pour des raisons de politique de confidentialité des données, nous avons été amenés à générer des données simulées pour les fichiers sources sauf le fichier Agence.xls qui contient les informations des agences de la banque. Ces fichiers sont les suivants :

-Nom_Prenom.xls : C'est un fichier Excel contenant les noms et les prénoms des différents clients de la banque.

-Email.xls : C'est un fichier Excel contenant les E-mails des différents clients de la banque.

-Tel.xls : C'est un fichier Excel contenant les numéros de téléphone des différents clients de la banque.

-CIN.xls : C'est un fichier Excel contenant les numéros de CIN des différents clients de la banque.

-Age.xls : C'est un fichier Excel contenant les dates de naissance des différents clients de la banque.

-Gestionnaires.xls : C'est un fichier Excel contenant les informations des gestionnaires de la banque.

-Segmentation.xls : C'est un fichier Excel contenant la segmentation des clients de la banque. On trouve les marchés, les professions, les segments NMR, les sous segments et les spécificités.

Ces fichiers ont été générés avec le langage de programmation Python afin de simuler des données exploitables pour la partie de reporting. (les figures de 16 à 21)

```

C:\> Users > ewali > OneDrive > Bureau > Stage_PFE > Nom_Prenom_fictive.py > {} openpyxl
1 import openpyxl
2 import random
3
4 noms = ["Gharbi", "Ben Ali", "Toumi", "Larbi", "Hammami", "Mansour", "Haddad", "Ammar", "Gabsi", "Bouhlel", "Mabrouk", "Cherif", "Abdeljelil", "Chakroun", "Khemiri", "Bou
5
6 prenoms = ["Ahmed", "Mohamed", "Ali", "Hassan", "Omar", "Khaled", "Mounir", "Kais", "Nizar", "Zied", "Nabil", "Majdi", "Samir", "Riad", "Khalil", "Sami", "Anis", "Saber"
7
8 def generer_nom_prenom():
9     nom = random.choice(noms)
10    prenom = random.choice(prenoms)
11    return [nom,prenom]
12
13 #ouverture de fichier Excel
14 wb = openpyxl.load_workbook('NOM_PRENOM.xlsx')
15
16 # Sélection de la feuille de calcul
17 sheet = wb['Nom_F']
18
19 # Remplissage des cellules
20 for row_num in range(2,2002):
21     sheet.cell(row=row_num, column=11).value = random.choice(["",generer_nom_prenom()[0]])
22     sheet.cell(row=row_num, column=12).value = random.choice(["",generer_nom_prenom()[1]])
23
24 # Sauvegarde du fichier
25 wb.save('NOM_PRENOM.xlsx')

```

Figure 16 : Simulation du fichier Nom_Prenom.xls

```

C:\> Users > ewali > OneDrive > Bureau > Stage_PFE > Email_fictive.py > {} openpyxl
1 import openpyxl
2 import random
3 import string
4
5 def generate_Email(Nom,Prenom):
6     """Génère un numéro de téléphone aléatoire"""
7     identifiant = ''.join(random.choice(string.ascii_lowercase + string.digits) for _ in range(6))
8     Email = str(Nom) + "." + str(Prenom) + identifiant + random.choice(["@gmail.com","@yahoo.fr","@hotmail.com","@outlook.com"])
9     return Email
10
11 #ouverture de fichier Excel
12 wb = openpyxl.load_workbook('EMAIL.xlsx')
13 NomPrenom = openpyxl.load_workbook('NOM_PRENOM.xlsx')
14
15 # Sélection de la feuille de calcul
16 sheet = wb['Email_F']
17 sheet_NP = NomPrenom['Nom_F']
18
19 # Remplissage des cellules
20 for row_num in range(2,2002):
21     nom = sheet_NP.cell(row=row_num, column=11).value
22     prenom = sheet_NP.cell(row=row_num, column=12).value
23     sheet.cell(row=row_num, column=11).value = generate_Email(nom,prenom)
24
25 # Sauvegarde du fichier
26 wb.save('EMAIL.xlsx')

```

Figure 17 : Simulation du fichier Email.xls

```

C:\> Users > ewali > OneDrive > Bureau > Stage_PFE > Tel_fictive.py > {} openpyxl
1 import openpyxl
2 import random
3
4
5 def generate_phone_number():
6     """Génère un numéro de téléphone aléatoire"""
7     phone_number = random.choice(["+216","216",""])
8     for i in range(8):
9         phone_number += str(random.randint(0, 9)) # Ajoute un chiffre aléatoire entre 0 et 9
10    phone_number += random.choice([" ",","])
11 return phone_number
12
13
14 #ouverture de fichier Excel
15 wb = openpyxl.load_workbook('TEL.xlsx')
16
17 # Sélection de la feuille de calcul
18 sheet = wb['Telephone_F']
19
20 # Remplissage des cellules
21 for row_num in range(2,2002):
22     sheet.cell(row=row_num, column=11).value = generate_phone_number()
23     sheet.cell(row=row_num, column=12).value = generate_phone_number()
24
25 # Sauvegarde du fichier
26 wb.save('TEL.xlsx')

```

Figure 18 : Simulation du fichier Tel.xls

```

C:\> Users > ewali > OneDrive > Bureau > Stage_PFE > CIN_fictive.py > {} openpyxl
1 import openpyxl
2 import random
3
4
5 def generate_CIN_number():
6     """Génère un numéro de CIN aléatoire"""
7     CIN_number = str(random.randint(0, 1))
8     for i in range(7):
9         CIN_number += str(random.randint(0, 9)) # Ajoute un chiffre aléatoire entre 0 et 9
10 return CIN_number
11
12
13 #ouverture de fichier Excel
14 wb = openpyxl.load_workbook('CIN.xlsx')
15
16 # Sélection de la feuille de calcul
17 sheet = wb['CIN_F']
18
19 # Remplissage des cellules
20 for row_num in range(2,2002):
21     sheet.cell(row=row_num, column=11).value = generate_CIN_number()
22
23 # Sauvegarde du fichier
24 wb.save('CIN.xlsx')

```

Figure 19 : Simulation du fichier CIN.xls

```

C:\> Users > ewali > OneDrive > Bureau > Stage_PFE > MPS_fictive.py > {} openpyxl
1 import openpyxl
2
3 # Charger le fichier Excel
4 wb = openpyxl.load_workbook('CLIENT.xlsx')
5
6 # Sélection de la feuille de calcul
7 sheet = wb['Client_F']
8 # Creation d'une colonne qui contient le marche,la profession,le segment
9
10 print(sheet.max_row)
11
12 Segmentation_liste = []
13 for row_num in range(2,2002) :
14     Segmentation_chaine = str(sheet.cell(row=row_num, column=2).value) +"," +str(sheet.cell(row=row_num, column=3).value)+"," +str(sheet.cell(row=row_num, column=4).value)+","
15     if Segmentation_chaine not in Segmentation_liste:
16         Segmentation_liste.append(Segmentation_chaine)
17
18 wb = openpyxl.load_workbook('Segmentation.xlsx')
19
20 # Sélection de la feuille de calcul
21 sheet = wb['Segmentation_F']
22
23 # Remplissage des cellules
24
25 for row_num in range(0,len(Segmentation_liste)) :
26     sheet.cell(row=row_num, column=1).value = row_num
27     sheet.cell(row=row_num, column=2).value = Segmentation_liste[row_num].split(",")[-1]
28     sheet.cell(row=row_num, column=3).value = Segmentation_liste[row_num].split(",")[-2]
29     sheet.cell(row=row_num, column=4).value = Segmentation_liste[row_num].split(",")[-3]
30     sheet.cell(row=row_num, column=5).value = Segmentation_liste[row_num].split(",")[-4]
31     sheet.cell(row=row_num, column=6).value = Segmentation_liste[row_num].split(",")[-5]
32
33 # Sauvegarde du fichier
34 wb.save('Segmentation.xlsx')

```

Figure 20 : Simulation du fichier Segmentation.xls

```

MPS_fictive.py  Gestionnaire_fictive.py
C:\Users\ewali>OneDrive>Bureau>Stage_PFE>Gestionnaire_fictive.py > {} pd
1 import pandas as pd
2 import openpyxl
3 import random
4
5
6 noms = ["Gharbi", "Ben Ali", "Toumi", "Larbi", "Hammami", "Mansour", "Haddad", "Ammar", "Gabsi", "Bouhlel", "Mabrouk", "Cherif", "Abdeljelil", "Chakroun", "Khemiri", "Bou
7
8 prenoms = ["Ahmed", "Mohamed", "Ali", "Hassan", "Omar", "Khaled", "Mounir", "Kais", "Nizar", "Zied", "Nabil", "Majdi", "Samir", "Riad", "Khalil", "Sami", "Anis", "Saber"
9
10 def generer_nom_prenom():
11     nom = random.choice(noms)
12     prenom = random.choice(prenoms)
13     return [nom,prenom]
14
15 # Charger le fichier Excel
16 df = pd.read_excel('CLIENT.xlsx')
17
18 # Supprimer les doublons et mettre les valeurs dans une liste
19 liste_valeurs = df['CODE_GESTIONNAIRE'].unique().tolist()
20
21 wb = openpyxl.load_workbook('GESTIONNAIRE.xlsx')
22
23 # Sélection de la feuille de calcul
24 sheet = wb['Gestionnaire_F']
25
26 # Remplissage des cellules
27
28 for row_num in range(0,len(liste_valeurs)) :
29     sheet.cell(row=row_num+2, column=1).value = liste_valeurs[row_num]
30     sheet.cell(row=row_num+2, column=2).value = generer_nom_prenom()[0]
31     sheet.cell(row=row_num+2, column=3).value = generer_nom_prenom()[1]
32
33 # Sauvegarde du fichier
34 wb.save('GESTIONNAIRE.xlsx')

```

Figure 21 : Simulation du fichier Gestionnaires.xls

2. Réalisation du processus ETL

Une étape cruciale de tout projet de business intelligence est la création du processus ETL après avoir rassemblé et préparé les données nécessaires pour l'entrepôt de données. Cette étape est essentielle pour garantir la qualité et la pertinence des données dans l'entrepôt, en les extrayant, en les transformant et en les chargeant dans le format adéquat.

2.1. Crédation des Jobs, importation des fichiers sources et configuration de la connexion à la base de données ORACLE

Dans cette section, nous allons expliquer comment nous avons mis en place notre environnement de travail en utilisant le logiciel Talend Open Studio for Data Integration.

1. Tout d'abord nous avons commencé par la création des jobs destinés au chargement de l'ODS (Operational Data Store) qui permet de stocker et centraliser des données provenant de sources différentes. La figure 22 montre les jobs que nous avons créés pour la partie extraction.



Figure 22 : Liste des jobs de l'ODS

2. Ensuite, dans l'onglet "métadonnées", nous avons importé les informations relatives aux fichiers Excel, qui ont été présentées dans la section " Génération de données simulées". La Figure 23 présente les métadonnées des fichiers sources importées.

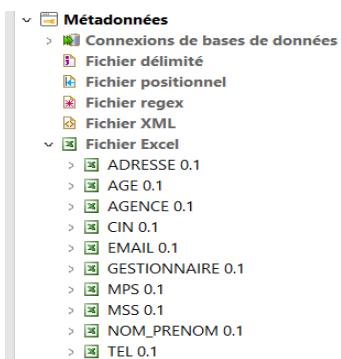


Figure 23 : Métadonnées des fichiers Excel

3. Enfin, nous procédons à la configuration de la connexion à la base de données de type ORACLE « Detect_Anomalies » (figure 24).

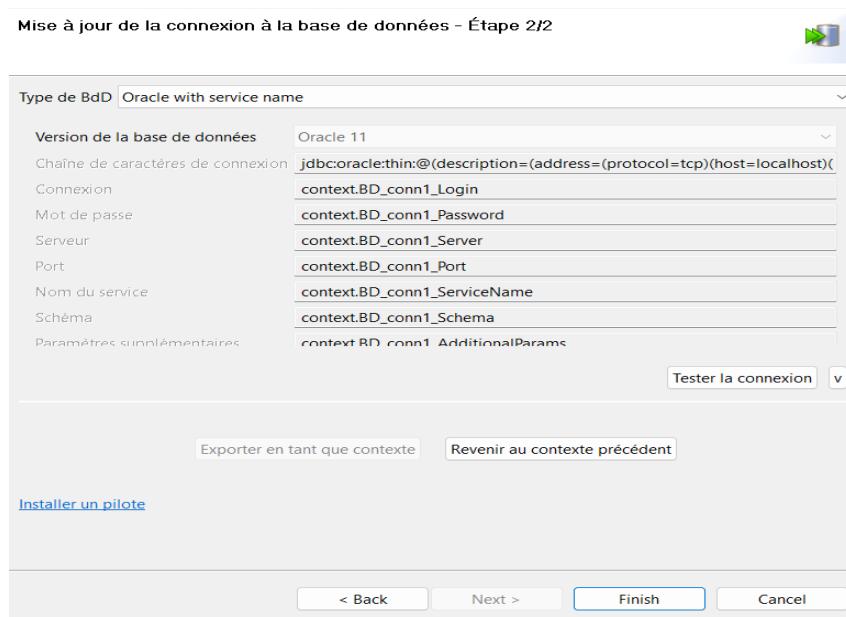


Figure 24 : Connexion à la base de données ORACLE

2.2. Chargement des fichiers sources dans la base de données ORACLE

Cette étape consiste à charger les données provenant des fichiers Excel dans les tables de la base « Detect_Anomalies ». Les figures de 25 à 38 montrent les jobs concernant le chargement de ces fichiers dans la base.

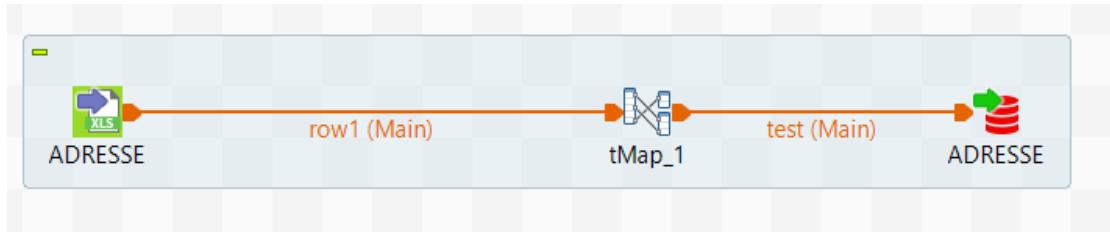


Figure 25 : Espace de travail du job « adresse »

Capture d'écran de l'interface Oracle Database. À gauche, l'explorateur de connexions montre une connexion à la base de données « Detect_Anomalies » avec une table nommée « ADRESSE ». À droite, une fenêtre de liste affiche les données de cette table. La table contient des colonnes telles que TE_OUVERTURE, DATE_MODIFICATION, CODE_GESTIONNAIRE, AGENCIE_CLIENT, ADRESSE, VILLE, REGION, PAYS, CODE_POSTAL, etc. Les données sont listées par ligne, avec des numéros de ligne et des dates d'ouverture et de modification.

Figure 26 : Chargement de la table « adresse »

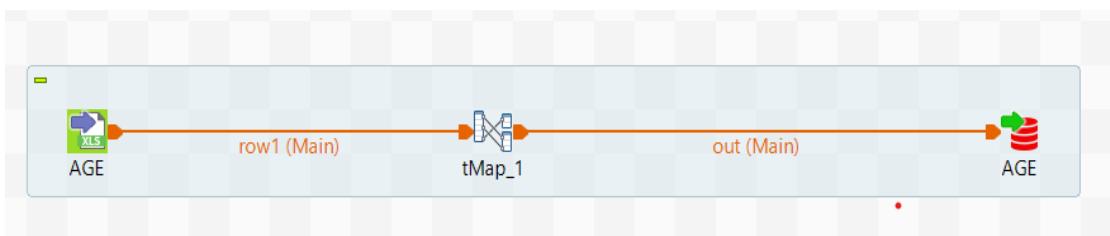


Figure 27 : Espace de travail du job « âge »

Connexions

script_MAJ_Mensuelle.sql

Page de bienvenue

Detect_Anomalies

AGE

Colonnes Données Modèle Contraintes Droits Statistiques Déclencheurs Flashback Dépendances Détails Partitions Index SQL Actions...

Trier... Filtrer :

ROME	PROFESSION	SEGMENT_NMR	SOUS_SEGMENT	SPECIFIQUE	DATE_OUVERTURE	DATE_MODIFICATION	CODE_GESTIONNAIRE	AGENCE_CLIENT	DATE_NAISANCE	
1	P14	S5	I4	TR GEI	10/06/15	20/07/15	612	16	Mon Jul 15 00:00:00 WAT 1996	
2	session liberale	P11	S3	I5	TR GEI	28/12/11	06/02/12	459	221	Fri Aug 17 00:00:00 WAT 1900
3	sprise	P17	S4	I5	VIP	24/10/15	03/12/15	102	190	Sat Dec 05 00:00:00 WAT 2009
4		P6	S2	I5	TR GEI	30/03/13	09/05/13	164	133	Mon Jul 01 00:00:00 WAT 1973
5		P8	S4	I4	BANQUE PRIVEE	29/12/14	07/02/15	204	49	Mon Jun 29 00:00:00 WAT 1998
6	session liberale	P11	S1	I5	OFF SHORE	31/01/19	12/03/19	986	111	Sun Feb 18 00:00:00 WAT 2007
7	iculiers	P2	S1	I4	VIP	02/10/17	11/11/17	611	30	m09/09/1966
8	iculiers	P1	S2	I3	BANQUE PRIVEE	12/08/08	21/09/08	898	23	Sun Jan 21 00:00:00 WAT 2018
9	sprise	P16	S3	I3	OFF SHORE	17/08/09	26/09/09	551	119	Tue May 24 00:00:00 WAT 1955
10	iculiers	P3	S1	I3	DIGITAL TSF	27/03/10	06/05/10	989	151	Fri Jan 30 00:00:00 WAT 1995
11	sprise	P16	S2	I4	DIGITAL TSF	25/05/11	04/07/11	545	40	Sat Oct 03 00:00:00 WAT 2015
12	sprise	P17	S5	I3	TR GEI	05/05/12	14/06/12	977	66	*21/12/1958
13		P12	S4	I3	TR GEI	22/06/17	01/08/17	258	18	Sat Mar 12 00:00:00 WAT 1994
14	sprise	P17	S2	I3	BANQUE PRIVEE	07/05/13	16/06/13	790	45	Sat Apr 14 00:00:00 WAT 1979
15	sprise	P17	S1	I1	VIP	18/04/11	08/04/23	867	130	Mon May 26 00:00:00 WAT 2014
16		P13	S1	I3	BANQUE PRIVEE	16/07/09	08/04/23	714	200	20/01/2007
17		P7	S2	I3	DIGITAL TSF	20/11/08	08/04/23	818	56	Wed May 07 00:00:00 WAT 1930
18	iculiers	P1	S5	I1	OFF SHORE	18/06/09	08/04/23	384	29	Wed Nov 05 00:00:00 WAT 1941
19	iculiers	P3	S3	I3	BANQUE DIGITALE	24/09/09	08/04/23	737	143	Tue Mar 30 00:00:00 WAT 1999
20	sprise	P17	S5	I4	VIP	04/04/17	08/04/23	478	94	Mon Apr 04 00:00:00 WAT 1977
21	session liberale	P10	S1	I2	DIGITAL TSF	25/05/13	08/04/23	377	185	Mon Jul 19 00:00:00 WAT 1999
22	sprise	P16	S1	I2	TR GEI	26/06/21	08/04/23	305	50	Sun Apr 11 00:00:00 WAT 1948
23		P8	S2	I4	OFF SHORE	02/10/18	08/04/23	118	233	Tue Dec 28 00:00:00 WAT 2010
24	iculiers	P3	S1	I3	DIGITAL TSF	15/01/14	08/04/23	368	46	Wed Aug 19 00:00:00 WAT 1953
25		P8	S4	I3	VIP	12/04/10	08/04/23	741	9	Sun Jul 24 00:00:00 WAT 1966
26	iculiers	P4	S5	I4	DIGITAL TSF	24/06/20	03/08/20	694	50	Mon Dec 28 00:00:00 WAT 1931
27	iculiers	P1	S2	I4	DIGITAL TSF	11/09/21	21/10/21	870	219	Sun Dec 23 00:00:00 WAT 1951
28	iculiers	ok	xx	xx	BANQUE DIGITALE	19/07/08	22/05/08	918	9	Sun Nov 08 00:00:00 WAT 1977

Figure 28 : Chargement de la table « âge »

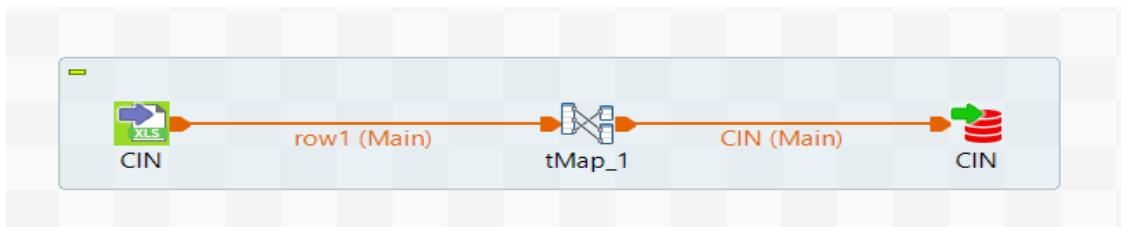


Figure 29 : Espace de travail du job « CIN »

Connexions

script_MAJ_Mensuelle.sql

Page de bienvenue

Detect_Anomalies

CIN

Colonnes Données Modèle Contraintes Droits Statistiques Déclencheurs Flashback Dépendances Détails Partitions Index SQL Actions...

Trier... Filtrer :

CODE_CLIENT	MARCHE	PROFESSION	SEGMENT_NMR	SOUS_SEGMENT	SPECIFIQUE	DATE_OUVERTURE	DATE_MODIFICATION	CODE_GESTIONNAIRE	AGENCE_CLIENT	NUM_CIN
1 1	TPE	P14	S5	I4	TR GEI	10/06/15	20/07/15	612	16	08561653
2 2	Profession liberale	P11	S3	I5	TR GEI	28/12/11	06/02/12	459	221	00566556
3 3	Entreprise	P17	S4	I5	VIP	24/10/15	03/12/15	102	190	06324083
4 4	TRE	P6	S2	I5	TR GEI	30/03/13	05/05/13	164	133	13639120
5 5	TRE	P8	S4	I4	BANQUE PRIVEE	29/12/14	07/02/15	204	49	01201545
6 6	Profession liberale	P11	S1	I5	OFF SHORE	31/01/19	12/03/19	986	111	04238559
7 7	Particuliers	P2	S1	I4	VIP	02/10/17	11/11/17	611	30	92718961
8 8	Particuliers	P1	S2	I3	BANQUE PRIVEE	12/08/08	21/09/08	898	23	47262344
9 9	Entreprise	P16	S3	I3	OFF SHORE	17/08/09	26/09/09	551	119	13011952
10 10	Particuliers	P3	S1	I3	DIGITAL TSF	27/03/10	06/05/10	989	151	06335079
11 11	Entreprise	P16	S2	I4	DIGITAL TSF	25/05/11	04/07/11	545	40	00746758
12 12	Entreprise	P17	S5	I3	TR GEI	05/05/12	14/06/12	977	66	10851809
13 13	TPE	P12	S4	I3	TR GEI	22/06/17	01/08/17	258	18	01117496
14 14	Entreprise	P17	S2	I3	BANQUE PRIVEE	07/05/13	16/06/13	790	45	15294921
15 15	Entreprise	P17	S1	I1	VIP	18/04/11	08/04/23	867	130	06190786
16 16	TPE	P13	S1	I3	BANQUE PRIVEE	16/07/09	08/04/23	714	200	18674501
17 17	TRE	P7	S2	I3	DIGITAL TSF	20/11/08	08/04/23	818	56	13066218
18 18	Particuliers	P1	S5	I1	OFF SHORE	18/06/09	08/04/23	384	29	15362620
19 19	Particuliers	P3	S3	I3	BANQUE DIGITALE	24/05/09	08/04/23	737	143	06399430
20 20	Entreprise	P17	S5	I4	VIP	04/04/17	08/04/23	478	94	11892245
21 21	Profession liberale	P10	S1	I2	DIGITAL TSF	25/05/13	08/04/23	377	185	15522346
22 22	Entreprise	P16	S1	I2	TR GEI	26/06/21	08/04/23	305	50	18799550
23 23	TRE	P8	S2	I4	OFF SHORE	02/10/18	08/04/23	118	233	19663236
24 24	Particuliers	P3	S1	I3	DIGITAL TSF	15/01/14	08/04/23	368	46	0367859
25 25	TRE	P8	S4	I3	VIP	12/04/10	08/04/23	741	9	02778119
26 26	Particuliers	P4	S5	I4	DIGITAL TSF	24/06/20	03/08/20	694	50	17503693
27 27	Particuliers	P1	S2	I4	DIGITAL TSF	11/09/21	21/10/21	870	219	10459080
28 28	iculiers	ok	xx	xx	BANQUE DIGITALE	19/07/08	22/05/08	918	9	06847377

Figure 30 : Chargement de la table « CIN »

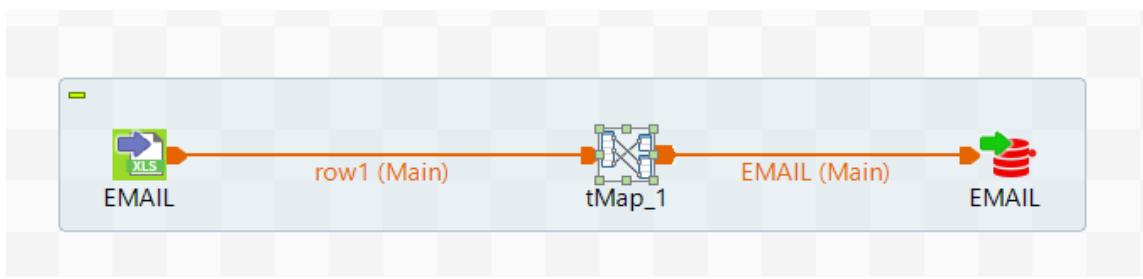


Figure 31 : Espace de travail du job « Email »

Figure 32 : Chargement de la table « EMAIL »

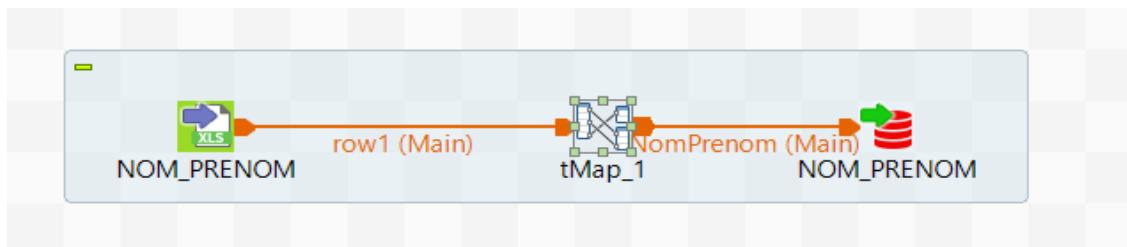


Figure 33 : Espace de travail du job « Noms et Prénoms »

CODE_CLIENT	MARCHE	PROFESSION	SEGMENT_NMR	SOUS_SEGMENT	SPECIFICITE	DATE_OUVERTURE	DATE_MODIFICATION	CODE_GESTIONNAIRE	AGENCE_CLIENT	NOM
1 1	TPE	P14	S5	I4	TR GEI	10/06/15	20/07/15	612	16	Hamed Mo
2 2	Profession libérale	P11	S3	I5	TR GEI	28/12/11	06/02/12	459	221	Chakroun Ri
3 3	Entreprise	P17	S4	I5	VIP	24/10/15	03/12/15	102	190	Cherif (n)
4 4	TRE	P6	S2	I5	TR GEI	30/03/13	09/05/13	164	133	(null) Sa
5 5	TRE	P8	S4	I4	BANQUE PRIVEE	29/12/14	07/02/15	204	49	Sabri (n)
6 6	Profession libérale	P11	S1	I5	OFF SHORE	31/01/19	12/03/19	986	111	Bouahmed (n)
7 7	Particuliers	P2	S1	I4	VIP	02/10/17	11/11/17	611	30	Mabrouk Ra
8 8	Particuliers	P1	S2	I3	BANQUE PRIVEE	12/08/08	21/09/08	898	23	Mabrouk (n)
9 9	Entreprise	P16	S3	I3	OFF SHORE	17/08/09	26/09/09	551	119	Klila (n)
10 10	Particuliers	P3	S1	I3	DIGITAL TSF	27/03/10	06/05/10	989	151	(null) (n)
11 11	Entreprise	P16	S2	I4	DIGITAL TSF	25/05/11	04/07/11	545	40	Boussaid (n)
12 12	Entreprise	P17	S5	I3	TR GEI	05/05/12	14/06/12	977	66	Boukaddida Ra
13 13	TPE	P12	S4	I3	TR GEI	22/06/17	01/08/17	258	18	Mansour Ya
14 14	Entreprise	P17	S2	I3	BANQUE PRIVEE	07/05/13	16/06/13	790	45	Gherbi Ri
15 15	Entreprise	P17	S1	I1	VIP	18/04/11	08/04/23	867	130	(null) Ra
16 16	TPE	P13	S1	I3	BANQUE PRIVEE	16/07/09	08/04/23	714	200	Toumi So
17 17	TRE	P7	S2	I3	DIGITAL TSF	20/11/09	08/04/23	818	56	(null) Sa
18 18	Particuliers	P1	S5	I1	OFF SHORE	18/06/09	08/04/23	384	29	(null) So
19 19	Particuliers	P3	S3	I3	BANQUE DIGITALE	24/09/09	08/04/23	737	143	Ammar Am
20 20	Entreprise	P17	S5	I4	VIP	04/04/17	08/04/23	478	94	Ben Ali (n)
21 21	Profession libérale	P10	S1	I2	DIGITAL TSF	25/05/13	08/04/23	377	185	(null) (n)
22 22	Entreprise	P16	S1	I2	TR GEI	26/06/21	08/04/23	305	50	(null) Ra
23 23	TRE	P8	S2	I4	OFF SHORE	02/10/18	08/04/23	118	233	(null) (n)
24 24	Particuliers	P3	S1	I3	DIGITAL TSF	15/01/14	08/04/23	368	46	(null) Im
25 25	TRE	P8	S4	I3	VIP	12/04/10	08/04/23	741	9	(null) (n)
26 26	Particuliers	P4	S5	I4	DIGITAL TSF	24/06/20	03/08/20	694	50	(null) Sa
27 27	Particuliers	P1	S2	I4	DIGITAL TSF	11/09/21	21/10/21	870	219	Toumi Sa
28 28	Particuliers	ex	ex	ex	BANQUE DUTRADE	17/03/08	22/03/08	318	?	(null) (n)

Figure 34 : Chargement de la table « Nom_Prenom »



Figure 35 : Espace de travail du job « Tel »

PROFESSION	SEGMENT_NMR	SOUS_SEGMENT	SPECIFICITE	DATE_OUVERTURE	DATE_MODIFICATION	CODE_GESTIONNAIRE	AGENCE_CLIENT	TELEPHONE_FIABLE	TELEPHONE_PORTABLE
1 P14	S5	I4	TR GEI	10/06/15	20/07/15	612	16	2161 6 0 7 8	2168 54 9 8
2 P11	S3	I5	TR GEI	28/12/11	06/02/12	459	221	2163 36 17 369	+21661552121
3 P17	S4	I5	VIP	24/10/15	03/12/15	102	190	21629 2 546 7	754 33 26 3
4 P6	S2	I5	TR GEI	30/03/13	09/05/13	164	133	+2166 4 8 7 2 62 5	2162324 7 52 0
5 P8	S4	I4	BANQUE PRIVEE	29/12/14	07/02/15	204	49	21608 71 1 3 88	4579 10 21 31
6 P11	S1	I5	OFF SHORE	31/01/19	12/03/19	986	111	21674 637 5 0 3	2166 10 21 559
7 P2	S1	I4	VIP	02/10/17	11/11/17	611	30	2165 64 0 141 0	2167262 31 1 4
8 P1	S2	I3	BANQUE PRIVEE	12/08/08	21/09/08	898	23	+216593 69 4 5 75	6 292 70 927
9 P16	S3	I3	OFF SHORE	17/08/09	26/09/09	551	119	+2167 7 56 673 2	5 5 5 77068
10 P3	S1	I3	DIGITAL TSF	27/03/10	06/05/10	969	151	83 04 82 78	6414 3180
11 P16	S2	I4	DIGITAL TSF	25/05/11	04/07/11	545	40	80258 8 0 1	84 6606 11
12 P17	S5	I3	TR GEI	05/05/12	14/06/12	977	66	5 0 4 1 5 11 8	66 3 04 743
13 P12	S4	I3	TR GEI	22/06/17	01/08/17	258	18	2165 2 3181 8 7	21625 8369 4 7Yassine
14 P17	S2	I3	BANQUE PRIVEE	07/05/13	16/06/13	790	45	+216108 71 8 4 6	2168 0 48 8 6 0 3
15 P17	S1	I1	VIP	18/04/11	08/04/23	867	130	3 3 69 61 31	5806 6 3 9 5
16 P13	S1	I3	BANQUE PRIVEE	16/07/09	08/04/23	714	200	3 3 6 03 42	5 46 13852
17 P7	S2	I3	DIGITAL TSF	20/11/09	08/04/23	818	56	+2169 02 2 5 23 8	+2165 81556 0 0
18 P1	S5	I1	OFF SHORE	18/06/09	08/04/23	384	29	Sofiene2162 66626 80 7 6 60 8 928	
19 P3	S3	I3	BANQUE DIGITALE	24/09/09	08/04/23	737	143	6 3 329 599	+2169 04 98 31 8
20 P17	S5	I4	VIP	04/04/17	08/04/23	478	94	8 28 97 1 78	2160 8 3 392 24
21 P10	S1	I2	DIGITAL TSF	25/05/13	08/04/23	377	185	2167 53 8 101 8	5 127 1 418
22 P16	S1	I2	TR GEI	26/06/21	08/04/23	305	50	2164 0 0163 45	+2162 5 566 2 81
23 P8	S2	I4	OFF SHORE	02/10/18	08/04/23	118	233	21604 6 08032	513 5 195 5
24 P3	S1	I3	DIGITAL TSF	15/01/14	08/04/23	368	46	+21604 39 64 9 6	216999926 07
25 P8	S4	I3	VIP	12/04/10	08/04/23	741	9	2166 1497 9 1 4	21697 275 5 61
26 P4	S5	I4	DIGITAL TSF	24/06/20	03/08/20	694	50	926 2 07 68	2167 7 0 2 2 34 9
27 P1	S2	I4	DIGITAL TSF	11/09/21	21/10/21	870	219	+2161 4 8 32057	60 677672
28 se	ex	ex	BANQUE DUTRADE	17/03/08	22/03/08	318	?	760 9 1 8 88	+2169 0 0 1 7

Figure 36 : Chargement de la table « Tel »

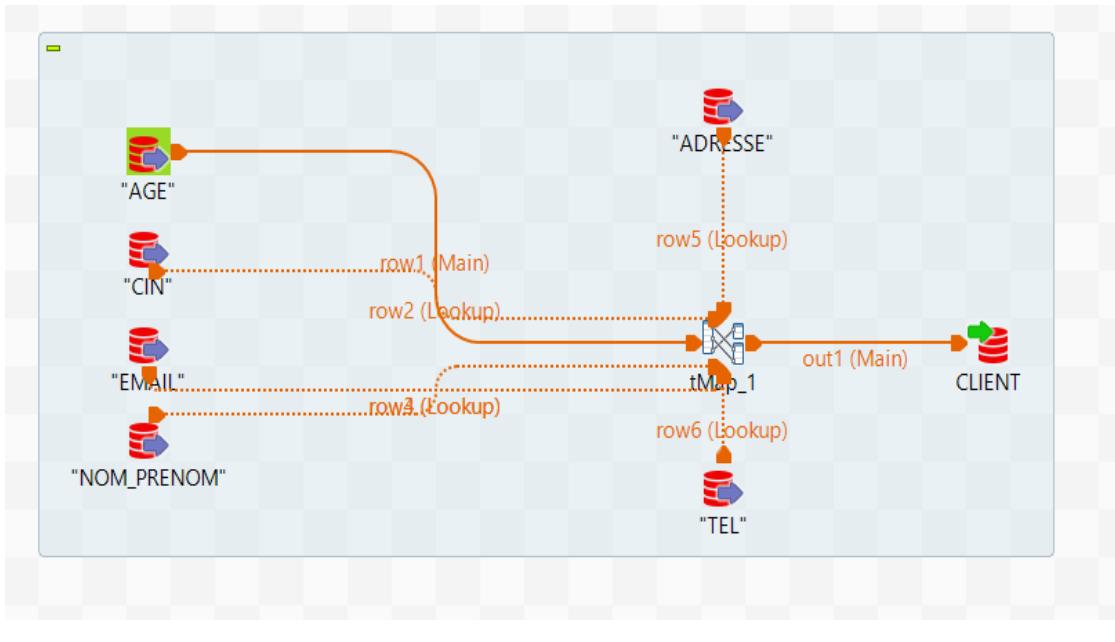


Figure 37 : Espace de travail du job « CLIENT »

Ce job (figure 37) permet de regrouper toutes les informations clientèles dans une seule table « CLIENT ». Cette table sera utile lors de la création de la table de fait.

CODE_CLIENT	NOM	PRENOM	NUM_CIN	EMAIL	TELEPHONE_FIABLE	TELEPHONE_PORTABLE	DATE_NAISSANCE	MARCHE	SEGMENT
1	1070	Ben Ali	Nour	65626546 Majri.Ahmedyygt@gmail.com	850 6 213 9	+21665 8 46 8 48	Sat Oct 12 00:00:00 WAT 1974 M1	S1	P3
2	1071	(null)	(null)	03393069 Soltani.Khalid2w8as@hotmail.fr	216167 45 3 9 9	+21688167 3 6	Sun May 11 00:00:00 WAT 1924 M1	S2	P5
3	1072	Abdeljelil	Majdi	10140552 Chakroun.Kaiselbervi@gmail.com	+21696 36 947 1	6 04 6 05 6 8	Fri Mar 18 00:00:00 WAT 2005 M4	S7	P21
4	1073	Khalfallah	(null)	13779142 Abid.Sofienepeyvad@hotmail.com	+216778 108 50	+2162 00 975 0 2	Sat Oct 26 00:00:00 WAT 1946 M3	S5	P13
5	1074	(null)	Yassmine	02333466 Cherni.Riadenestdus@hotmail.com	3 307 3 9 88	2163 30 7 4 3 0 8	Tue Sep 07 00:00:00 WAT 2010 M1	S2	P6
6	1075	(null)	17496127	Abdeljelil.Hassen13live@hotmail.fr	+2162 8 44 3 0 15	+2162 1 244 341	Tue Jun 15 00:00:00 WAT 1965 M5	S10	P28
7	1076	(null)	00669368	Cherif.Hassancoude@gmail.com	+2164 7 41 76 9 0	2164 3 64 54 69	Thu Oct 27 00:00:00 WAT 2016 M5	S10	P30
8	1077	Khemiri	Sami	15930387 Ben Ali.Onsaavdw@yahoo.fr	+2164 7 41 76 9 0	21620 34 78 2 2	Sun May 10 00:00:00 WAT 1942 M5	S9	P27
9	1078	Sahli	Nour	09442455 Abdeljelil.Nabilb75qd@gmail.com	2163 52 6 1 51	+21699 97 6 7 9	Thu Dec 31 00:00:00 WAT 2015 M2	S3	P7
10	1079	Hammami	Omar	03552550 Cherni.Aniszxkbtj@outlook.com	2164 6 1 39 44 6	2162 110 1 804	Tue Mar 17 00:00:00 WAT 2020 M1	S2	P5
11	1080	(null)	(null)	02288001 Mansouri.Samirvrxv0c@hotmail.com	+2162 7 0 22 54	+2165 1 5 17 09 7	Wed Sep 27 00:00:00 WAT 1961 M5	S9	P25
12	1081	(null)	10454596	Gharbi.MarwaVb5x8@hotmail.fr	2165 6 915 489	2 4 3 50 0 0	Tue Oct 13 00:00:00 WAT 2015 M1	S2	P5
13	1082	Bouhlel	Yassmine	18925755 Majri.Kais4elrgp@yahoo.fr	+2165 98 05 27 3	03656 56 3	Mon Feb 02 00:00:00 WAT 1931 M4	S8	P22
14	1083	(null)	Maissa	13504606 Gabsi.Sofiene8m3ru@hotmail.fr	3 03 0026 1	5 739 51 36	Mon Jan 05 00:00:00 WAT 1953 M2	S4	P10
15	1084	Ahdid	(null)	17075650 Soltani.Mounirrvhmk@outlook.com	+2164 616 3 3 56	216359723 5 2	Sun Jul 28 00:00:00 WAT 1968 M5	S9	P25
16	1085	Khemiri	Mourad	19409595 Mansouri.Rahma8m93e@hotmail.fr	+2166 1439 30 9	+2162 3375 9 54	Sun Jul 23 00:00:00 WAT 1961 M2	S4	P10
17	1086	(null)	(null)	06800258 Mansouri.Ahmedolbpf2@hotmail.com	5 04 08 584	91 6 98 40	Tue Jun 26 00:00:00 WAT 1934 M3	S6	P18
18	1087	Khemiri	Rahma	13497551 Cherif.Omar93p3rb@hotmail.com	+21646 8 73 9 2 3	21697 1 8 30 62	Tue Jul 01 00:00:00 WAT 1969 M3	S5	P15
19	1088	(null)	(null)	14756270 Cherni.Khaledwpob4@hotmail.fr	21624 217 247	+21600137 6 10	Wed Jul 04 00:00:00 WAT 1984 M5	S9	P26
20	1089	(null)	Omar	03822566 Khemiri.Omarou06@gmail.com	3 5 6 9 0 5 15	+21699 365 99 7	Wed Apr 29 00:00:00 WAT 1931 M1	S2	P4
21	1090	(null)	(null)	00751363 Boukadda.Kaisenhh2p@hotmail.fr	+2164 98 0 73 41	4 242821 6	Tue Jul 23 00:00:00 WAT 1946 M5	S9	P26
22	1091	Toumi	Zied	10782318 Cherni.AnisZhp707@hotmail.com	+21669 7 34 37	2161 815 1 1 9 3	Sat Apr 21 00:00:00 WAT 1951 M2	S3	P8
23	1092	(null)	(null)	19738598 Toumi.Mounir8al14@hotmail.fr	216452 1 4 7 4 1	21612 0284 88	Mon Sep 20 00:00:00 WAT 1971 M4	S8	P22
24	1093	(null)	Anis	18679282 Berrikche.Mounir00ury@gmail.com	2163 7 6 852 1 8	2 60 698 83	Tue Oct 16 00:00:00 WAT 1934 M5	S9	P26
25	1094	(null)	Thouray	14714247 Cherni.Hassan8m2m8@hotmail.fr	+216316 4 372 6	21608 0 4 63 35	Sun Oct 30 00:00:00 WAT 2023 M3	S6	P17
26	1095	(null)	Anis	09319550 Khalfallah.Nour1x7nf@outlook.com	1 5 3 6 2228	1 0 3 4 4 0 8 4	Sat Dec 17 00:00:00 WAT 2022 M4	S8	P24
27	1096	Ben Ali	Nabil	16346352 Chakroun.Samirvrxv0c@hotmail.com	2167 045 9366	216924 3 7 24 5	Wed Nov 20 00:00:00 WAT 1963 M1	S1	P3
28	1097	Ben Ali	Enfiane	01416797 Nabil.Matidi8m8@gmail.com	21611 4 17 476	21629 6 66 16 32	Wed Jan 08 00:00:00 WAT 1971 M3	S9	P9

Figure 38 : Chargement de la table « Client »

2.3. Application des règles de gestion pour transformer les données

Talend est l'outil idéal pour appliquer les règles de gestion. Deux tables seront comme output pour chaque table mise en input.

Le premier output correspond aux données incorrectes (anomalies détectées) et le deuxième est celui des données correctes.

Nous avons appliqué un ensemble de bonnes pratiques [5] pour l'utilisation des jobs dans Talend Open Studio for Data Integration. Nous avons pris soin de suivre ces pratiques pour assurer une bonne qualité de notre travail :

- Nommer les composants d'entrée et de sortie et mettre des titres dans chaque job.
- Utilisation de variable de contexte pour les métadonnées et la base pour faciliter leurs mises à jour. La figure 39 montre les contextes créés.



Figure 39 : Ensemble des contextes des métadonnées et de la base de données ORACLE « Detect_Anomalies »

- Utilisation de 3 jobs masters pour regrouper les jobs de chaque partie (un job master pour l'ODS, un autre pour les jobs concernant les transformations et un dernier pour les jobs des dimensions et tables de faits)

- Veiller à orchestrer les jobs masters en utilisant les composants **tPrejob** et **tPostjob** pour ouvrir la connexion à la base de données et la fermer une fois que le chargement a été effectué

- Veiller à bien nommer et organiser les dossiers contenant les jobs

2.3.1 Application des règles de gestion pour transformer les données liées aux adresses des clients

La figure 40 montre l'environnement du job concernant les transformations effectuées sur les données importées de la table « Adresse » qui a été chargée dans l'étape précédente.

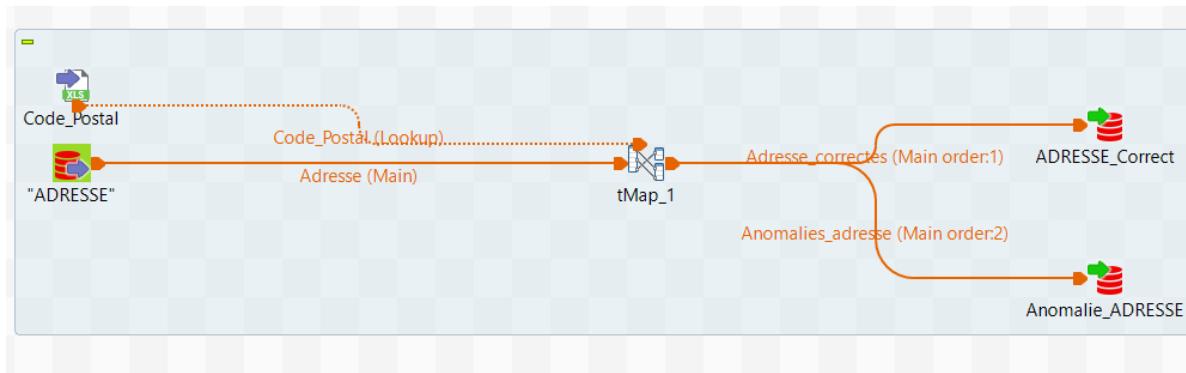


Figure 40 : Environnement de travail du job « T_Adresse »

Ce job permet d'appliquer les règles de gestion pour les adresses des clients. Une adresse est considérée comme adresse correcte si elle :

- Contient au moins un des mots suivants (route, bloc, avenue, av, rue, résidence, appartement, etc...).
- Contient au moins 3 mots.
- Contient au moins un chiffre.
- Contient au moins 2 lettres pour chaque mot.

1. La première étape pour la réalisation de ce job consiste à importer la table adresse et la métadonnée du fichier Excel « code_postal.xls » en utilisant respectivement les composants d'entrée **tOracleInput** et **tFileInputExcel** pour lire la table et le fichier. Les figures de 41 à 44 montrent ces composants ainsi que leur schéma respectif.



Figure 41 : Table « Adresse »

Schéma de "ADRESSE"

"ADRESSE"

Colonne utilisée	Colonne	Db Column	Clé	Type	Type ...	N.	Modèle d...	Long...	Précis...	Par...	Com...
CODE_CLIENT	CODE_CLIENT	CODE_CLIENT		String	VARC...			4	0		
MARCHE	MARCHE	MARCHE		String	VARC...			20	0		
PROFESSION	PROFESSION	PROFESSION		String	VARC...			20	0		
SEGMENT_NMR	SEGMENT_NMR	SEGMENT_NMR		String	VARC...			20	0		
SOUS_SEGME...	SOUS_SEGMENT	SOUS_SEGMENT		String	VARC...			20	0		
SPECIFICITE	SPECIFICITE	SPECIFICITE		String	VARC...			20	0		
DATE_OUVERTURE	DATE_OUVERTURE	DATE_OUVERTURE		Date	DATE		"dd/MM/...	7	0		
DATE_MODIFICA...	DATE_MODIFICA...	DATE_MODIFICA...		Date	DATE		"dd/MM/...	7	0		
CODE_GESTIO...	CODE_GESTIONN...	CODE_GESTIONN...		String	VARC...			3	0		
AGENCE_CLIE...	AGENCE_CLIENT	AGENCE_CLIENT		String	VARC...			3	0		
ADRESSE	ADRESSE	ADRESSE		String	VARC...			100	0		
VILLE	VILLE	VILLE		String	VARC...			30	0		
REGION	REGION	REGION		String	VARC...			30	0		
PAYS	PAYS	PAYS		String	VARC...			30	0		
CODE_POSTAL	CODE_POSTAL	CODE_POSTAL		String	VARC...			4	0		

OK Cancel

Figure 42 : Structure de la table « Adresse »

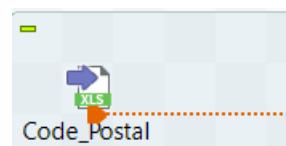


Figure 43 : Fichier « Code_Postal.xls »

Description du schéma

Colonne	Clé	Type	N.	Modèle de date ...	Longueur	Précision	Par dé...	Commen...
Localite		String			23	0		
Code_postal		Integer			4	0		
Delegation		String			19	0		

Figure 44 : Structure de « Code_Postal.xls »

2. La deuxième étape consiste à l'utilisation du composant tMap pour appliquer les règles de gestion afin d'avoir deux sorties : une table pour les informations qui contiennent des anomalies pour les adresses et une autre pour les adresses correctes. La figure 45 montre le composant tMap.

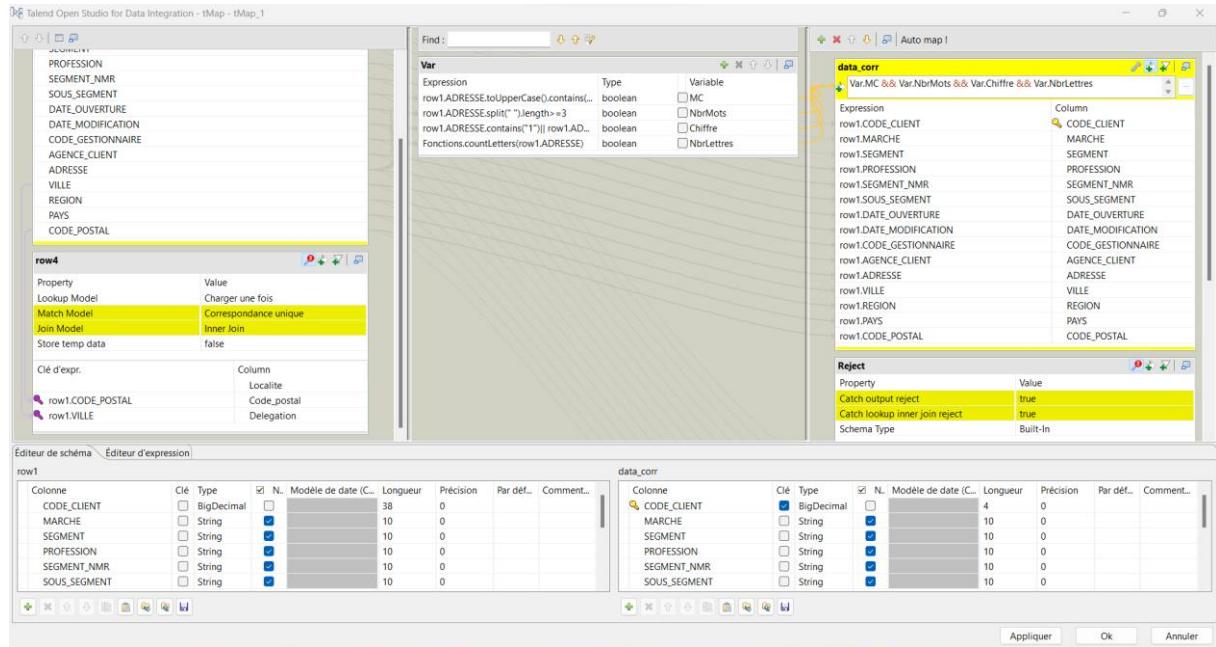


Figure 45 : Composant « tMap_Adresse »

La figure 45 présente l'interface de l'éditeur de mapping du composant tMap. À gauche, nous pouvons voir les schémas des tables d'entrée nommées « Adresse » et « Code_postal ». À droite, le schéma des tables de sortie « Adresses_correctes » et « Anomalies_Adresses » est affiché. Pour extraire les données de ces tables d'entrée, il est nécessaire de créer des jointures entre les tables qui partagent une même colonne. Pour ce faire, il suffit de faire glisser les clés de relation de la table primaire (Adresse) vers la table secondaire (Code_postal). Le type de jointure utilisé ici est une jointure interne. La figure 46 illustre les jointures effectuées.

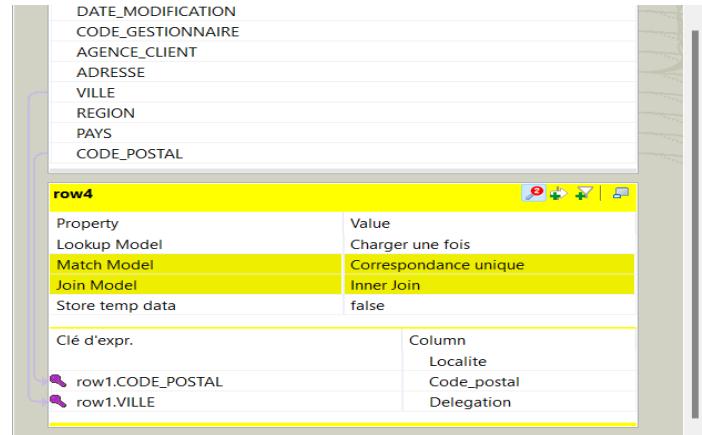


Figure 46 : Jointure des tables d'entrées

Une fois que les jointures sont réalisées, nous pouvons créer des variables pour appliquer les règles de gestion. Pour ce faire, nous allons utiliser le constructeur d'expression, qui nous permet de créer des variables à partir d'expressions complexes en utilisant des fonctions incluses dans Talend ou créées par les utilisateurs. Les figures de 47 à 51 présentent les expressions utilisées pour appliquer les règles de gestion.

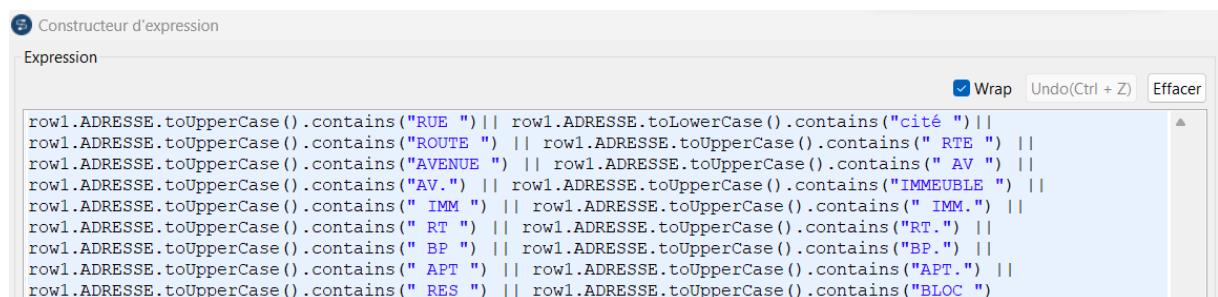


Figure 47 : Expression pour la règle de gestion qui consiste à avoir au moins un mot signifiant une adresse

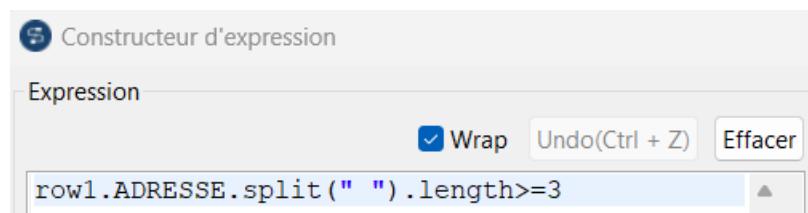


Figure 48 : Expression pour la règle de gestion qui consiste à avoir au moins 3 mots dans les adresses

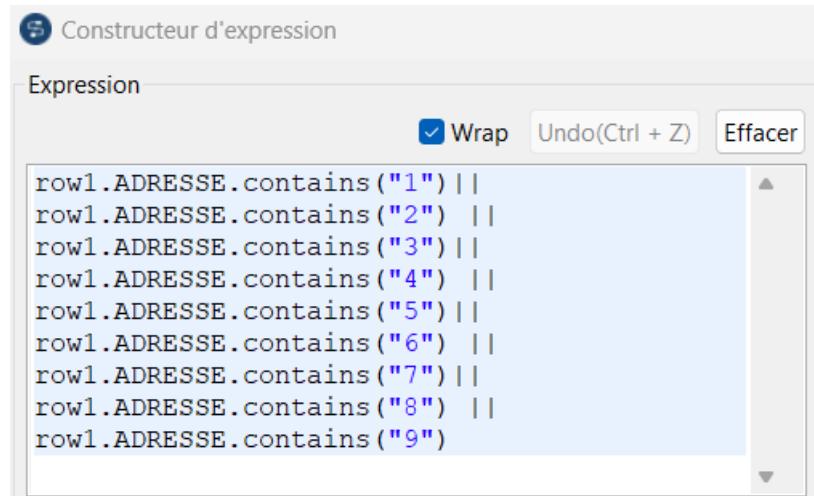


Figure 49 : Expression pour la règle de gestion qui consiste à avoir au moins un chiffre dans les adresses

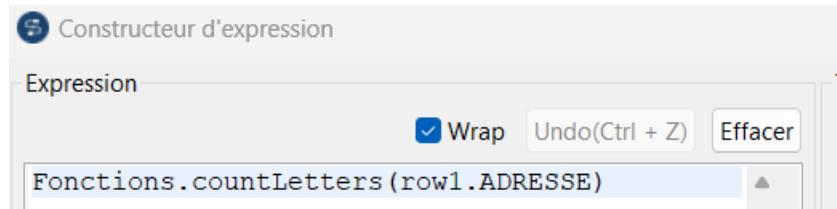


Figure 50 : Expression pour la règle de gestion qui consiste à avoir au moins deux lettres dans chaque mot

```
package routines;
import java.text.SimpleDateFormat;

public class Fonctions {

    public static boolean countLetters(String input) {
        String[] words = input.split(" "); // sépare la chaîne en mots

        for (int i = 0; i < words.length; i++) {
            String word = words[i];
            boolean isNumber = true;
            for (char c : word.toCharArray()) { // vérifie si chaque caractère du mot est un chiffre
                if (c < '0' || c > '9') {
                    isNumber = false;
                    break;
                }
            }
            if (!word.equals("N") && word.length() < 2 && !isNumber && !word.equals("N")) { // vérifie si le mot a moins de 2 lettres et n'est pas "N"
                return false; // si le mot ne satisfait pas les critères, renvoie false immédiatement
            }
        }
        return true; // si tous les mots ont 2 lettres ou plus (sauf "N" et les chiffres), renvoie true
    }
}
```

Figure 51 : Traitement de la fonction « countLetters »

Maintenant nous pouvons récupérer les colonnes dont nous avons besoin en les faisant glisser depuis les tables d'entrée vers les tables de sortie nommées « Adresses_correctes » et « Anomalies_Adresses » qui va contenir les adresses qui présentent des anomalies. Une fois les tables de sortie créées, nous pouvons fermer l'éditeur de mappage en cliquant sur OK, car nous

avons tout ce dont nous avons besoin. Il convient de noter qu'une visualisation des deux tables de sortie, « Adresses_correctes » et « Anomalies_Adresses », est présentée dans les figures 52 et 53 respectivement.

Table_Adresse_correctes	
Var.MC && Var.NbrMots && Var.Chiffre && Var.NbrLettres	
Expression	Column
Table_Adresse.CODE_CLIENT	CODE_CLIENT
Table_Adresse.MARCHE	MARCHE
Table_Adresse.PROFESSION	PROFESSION
Table_Adresse.SEGMENT_NMR	SEGMENT_NMR
Table_Adresse.SOUS_SEGMENT	SOUS_SEGMENT
Table_Adresse.SPECIFICITE	SPECIFICITE
Table_Adresse.DATE_OUVERTURE	DATE_OUVERTURE
Table_Adresse.DATE_MODIFICATION	DATE_MODIFICATION
Table_Adresse.CODE_GESTIONNAIRE	CODE_GESTIONNAIRE
Table_Adresse.AGENCE_CLIENT	AGENCE_CLIENT
Table_Adresse.ADRESSE	ADRESSE
Table_Adresse.VILLE	VILLE
Table_Adresse.REGION	REGION
Table_Adresse.PAYS	PAYS
Table_Adresse.CODE_POSTAL	CODE_POSTAL

Figure 52 : Table « Adresse_correctes »

Table_Anomalies_adresse	
Property	Value
Catch output reject	true
Catch lookup inner join reject	true
Schema Type	Built-In
Expression	Column
Table_Adresse.CODE_CLIENT	CODE_CLIENT
Table_Adresse.MARCHE	MARCHE
Table_Adresse.PROFESSION	PROFESSION
Table_Adresse.SEGMENT_NMR	SEGMENT_NMR
Table_Adresse.SOUS_SEGMENT	SOUS_SEGMENT
Table_Adresse.SPECIFICITE	SPECIFICITE
Table_Adresse.DATE_OUVERTURE	DATE_OUVERTURE
Table_Adresse.DATE_MODIFICATION	DATE_MODIFICATION
Table_Adresse.CODE_GESTIONNAIRE	CODE_GESTIONNAIRE
Table_Adresse.AGENCE_CLIENT	AGENCE_CLIENT
Table_Adresse.ADRESSE	ADRESSE
Table_Adresse.VILLE	VILLE
Table_Adresse.REGION	REGION
Table_Adresse.PAYS	PAYS
Table_Adresse.CODE_POSTAL	CODE_POSTAL

Figure 53 : Table « Anomalies_adresse »

La jointure entre les tables d'entrée va nous servir à avoir les deux tables en sorties "Adresses_correctes" et "Anomalies_Adresses" en mettant la valeur de la propriété « Catch output reject » sur 'true' pour avoir les adresses qui ne respectent pas les règles de gestions et la valeur de la propriété « Catch lookup inner join reject » sur 'true' pour avoir les adresses dont les codes postaux ne sont pas corrects et ne correspondent pas à ceux qui se trouvent dans le fichier « code_postal.xls »

3. La troisième étape consiste à charger les tables dans la base de données ORACLE. Pour se faire nous avons utilisé le composant **tOracleOutput**. Les figures 54 et 55 montrent la configuration des deux composants **tOracleOutput** pour chaque table.

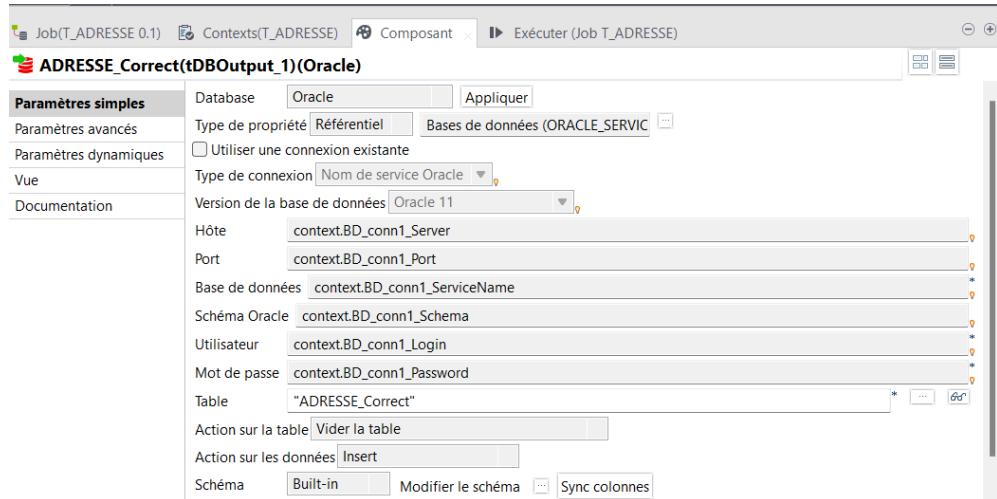


Figure 54 : Composant *tOracleOutput* pour la table « Adresse_Correctes »

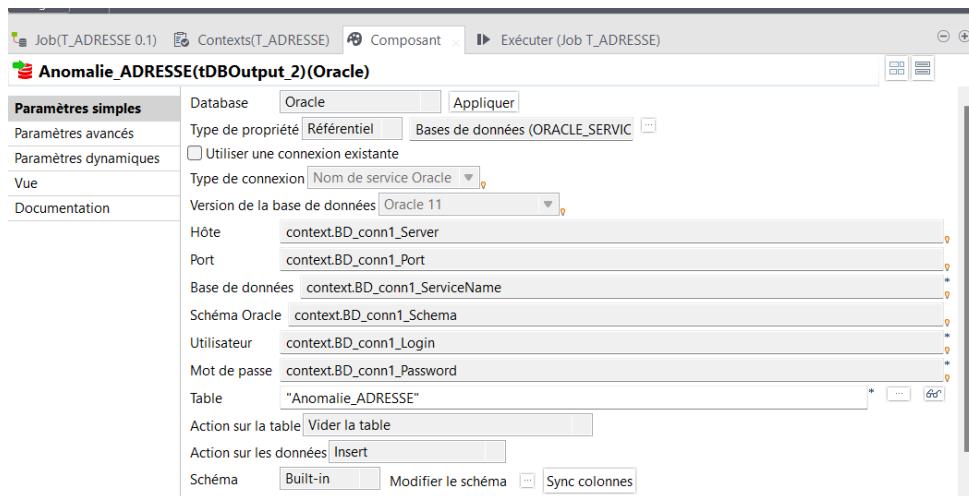


Figure 55 : Composant *tOracleOutput* pour la table « Anomalies_Adresse »

Les figures de 56 à 58 montrent l'exécution du Job « T_Adresse » (figure 56) ainsi que le chargement des deux tables "Adresses_correctes" et "Anomalies_Adresses" dans la base ORACLE (figure 57 et 58).

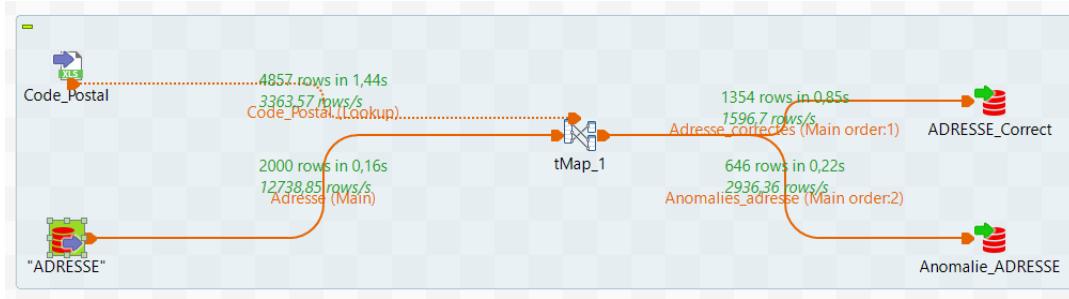


Figure 56 : Exécution du job « T_adresse »

Connexions

Page de terminaison | Accès direct | Aide en ligne | ANTHONIE_12_ADDRESS

Colonnes | Dossiers | Mode | Historique | Données | Statistiques | Débordements | Dépendances | Offtales | Partitions | Index | SQL

Actions | Annuler | Sauvegarder | Imprimer | Trier | Filtrer | Rechercher |

AUTOREGISTREATION | CODE_GESTIONNAIRE | AGENCE_CLIENT | ADRESSE | GESTIONNAIRE | CODE_POSTAL |

1 23/11/14 316 10/11/14 Eriahm Zone C 2031 Oued Elill Manouba Oued Elill Manouba Tunisie 1063

2 02/05/12 555 10/11/14 Eriahm F 2031 Oued Elill Manouba Oued Elill Manouba Tunisie 2031

3 25/04/15 603 11/08/14 El Ghabette 7020 Sidi Sedjane Bizerte Sidi Sedjane Tunisie 1063

4 24/02/14 421 19/07/14 El Ghazala 6000 March Gabes March Gabes Tunisie 1063

5 31/07/14 457 13/11/14 El Ghazedi 6000 March Gabes March Gabes Tunisie 1063

6 07/01/11 676 44/11/14 El Ghazedi 6000 March Gabes March Gabes Tunisie 2032

7 12/11/14 967 10/11/14 El Ghazali 1 2003 Raedess Ariana Recaud Ariana Tunisie 1063

8 23/01/14 797 10/11/14 El Ghazali 1 2003 Raedess Ariana Recaud Ariana Tunisie 1063

9 05/12/10 434 7/9/11 El Ghoubare 4041 Soueide Jendouze Soueide Jendouze Tunisie 2045

10 01/01/16 313 1/11/16 El Ghribi 2031 Sidi Thabet Ariana Sidi Thabet Ariana Tunisie 1063

11 01/01/16 347 19/07/14 El Ghribi 2031 Sidi Thabet Ariana Sidi Thabet Ariana Tunisie 1063

12 01/01/16 318 9/11/14 El Gias 9401 Ben Arous Beja Tebourba Tunisie 1063

13 9 15/01/16 116 11/10/14 El Gouiaha 2084 El Mhilia Ariana El Mhilia Ariana Tunisie 1075

14 5/05/11/16 396 3/10/16 El Gouiaha 2020 Sidi Thabet Ariana Sidi Thabet Ariana Tunisie 2079

15 23/01/14 714 10/11/14 El Hafsi 2000 Sidi Thabet Ariana Sidi Thabet Ariana Tunisie 1063

16 3 04/04/13 192 18/11/14 El Hane 145 Mohamedia Ben Arous Mohamedia Ben Arous Tunisie 1063

17 1 04/05/11 211 35/09/16 El Hane 4000 Sousse Medina Sousse Medina Tunisie 1063

18 27/01/14 200 10/11/14 El Hafez 2000 Sousse Medina Sousse Medina Tunisie 1063

19 22/04/12 815 17/9/11 El Hafez 2000 Sahloul Sahloul Sahloul Tunisie 2051

20 1 25/02/11 804 12/11/14 El Hala 1150 Idris Zaghouan Idris Zaghouan Tunisie 1063

21 3 11/02/12 231 17/9/11 El Houl 2045 Ben Arous Ben Arous Tunisie 1023

22 1 22/01/12 817 12/11/14 El Housni 2000 Ben Arous Ben Arous Tunisie 1063

23 9 01/04/16 764 32/11/14 El Idirbi 4224 Dous Nabeul Dous Nabeul Tunisie 1062

24 4 26/01/15 357 12/11/14 El Idirbi 8045 El Haouaria Sahloul El Haouaria Sahloul Tunisie 1062

25 9 08/01/20 504 25/11/14 El Idirbi 8046 Haouaria Sahloul Haouaria Sahloul Tunisie 1023

26 1 01/02/15 734 27/02/14 El Idirbi 8046 Haouaria Sahloul Haouaria Sahloul Tunisie 1062

27 5 10/01/16 672 30/01/16 El Idirbi Beyrou 2021 Dous Elill Manouba Oued Elill Manouba Tunisie 1063

28 1 04/01/16 744 30/01/16 El Idirbi Beyrou 2021 Dous Elill Manouba Oued Elill Manouba Tunisie 1063

Messages - Journal

Messages | Page Journalisation | Instructions

Figure 57 : Chargement de la table « Anomalies_Adresses »

Figure 58 : Chargement de la table « Adresses_correctes »

2.3.2 Application des règles de gestion pour transformer les données liées aux dates de naissance des clients

La figure 59 montre l'environnement du job concernant les transformations effectuées sur les données importées de la table « Age » qui a été chargée dans l'étape précédente.

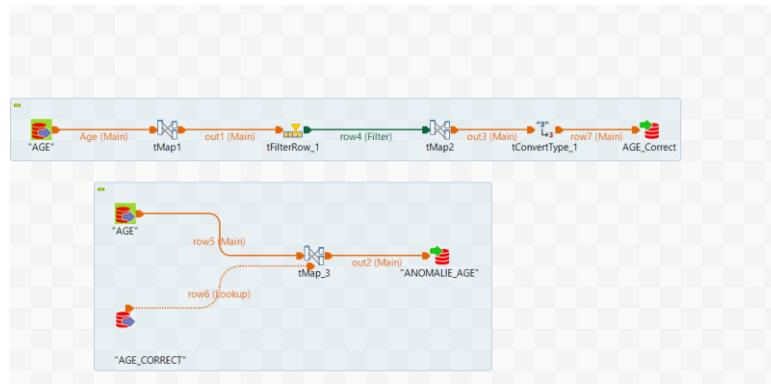


Figure 59 : Environnement de travail du job « T_Age »

Ce job permet d'appliquer les règles de gestion pour les adresses des clients. Une adresse est considérée comme date de naissance correcte si elle :

- Est sous la forme dd/mm/yyyy
- Entre 0 et 100 ans

1. La première étape pour la réalisation de ce job consiste à importer la table âge en utilisant le composant d'entrée **tOracleInput** pour lire la table et le fichier. Les figure 60 et 61 montrent ce composant ainsi que son schéma.

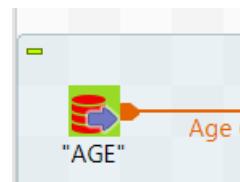


Figure 60 : Table « Age »

Nom	Age																																																																																																																							
Commentaire																																																																																																																								
Type	TABLE																																																																																																																							
Accès sur la table : AGE.																																																																																																																								
Récupérer le schéma																																																																																																																								
Utilisez le bouton "Récupérer le schéma" pour remplacer le schéma courant par le schéma de l'																																																																																																																								
Schéma																																																																																																																								
<table border="1"> <thead> <tr> <th>Colonne</th> <th>De Column</th> <th>Créé</th> <th>Type de BiO</th> <th>Type</th> <th>Modèle de date (Créé)</th> <th>Longueur</th> <th>Precision</th> <th>Par défaut</th> <th>Commentaire</th> </tr> </thead> <tbody> <tr> <td>CODE_CLIENT</td> <td>CODE_CLIENT</td> <td></td> <td><input type="checkbox"/> NUMBER</td> <td>BigDecimal</td> <td></td> <td>20</td> <td>0</td> <td></td> <td></td> </tr> <tr> <td>MARCHE</td> <td>MARCHE</td> <td></td> <td><input type="checkbox"/> VARCHAR(40)</td> <td>String</td> <td></td> <td>20</td> <td>0</td> <td></td> <td></td> </tr> <tr> <td>SECOMENT</td> <td>SECOMENT</td> <td></td> <td><input type="checkbox"/> VARCHAR(40)</td> <td>String</td> <td></td> <td>20</td> <td>0</td> <td></td> <td></td> </tr> <tr> <td>PROFESSION</td> <td>PROFESSION</td> <td></td> <td><input type="checkbox"/> VARCHAR(40)</td> <td>String</td> <td></td> <td>20</td> <td>0</td> <td></td> <td></td> </tr> <tr> <td>VILLECLIENT</td> <td>VILLECLIENT</td> <td></td> <td><input type="checkbox"/> VARCHAR(40)</td> <td>String</td> <td></td> <td>20</td> <td>0</td> <td></td> <td></td> </tr> <tr> <td>SOUS_SECOMENT</td> <td>SOUS_SECOMENT</td> <td></td> <td><input type="checkbox"/> VARCHAR(40)</td> <td>String</td> <td></td> <td>20</td> <td>0</td> <td></td> <td></td> </tr> <tr> <td>DATE_OUVERTURE</td> <td>DATE_OUVERTURE</td> <td></td> <td><input type="checkbox"/> DATE</td> <td>Date</td> <td></td> <td>20</td> <td>0</td> <td></td> <td></td> </tr> <tr> <td>DATE_MODIFICATION</td> <td>DATE_MODIFICATION</td> <td></td> <td><input type="checkbox"/> DATE</td> <td>Date</td> <td></td> <td>20</td> <td>0</td> <td></td> <td></td> </tr> <tr> <td>CODE_GEOGRAPHIQUE</td> <td>CODE_GEOGRAPHIQUE</td> <td></td> <td><input type="checkbox"/> NUMBER</td> <td>BigDecimal</td> <td>"dd/MM/yyyy"</td> <td>20</td> <td>0</td> <td></td> <td></td> </tr> <tr> <td>ADRESSE_CLIENT</td> <td>ADRESSE_CLIENT</td> <td></td> <td><input type="checkbox"/> NUMBER</td> <td>BigNumber</td> <td>"dd/MM/yyyy"</td> <td>20</td> <td>0</td> <td></td> <td></td> </tr> <tr> <td>DATE_NAISSANCE</td> <td>DATE_NAISSANCE</td> <td></td> <td><input type="checkbox"/> VARCHAR(40)</td> <td>String</td> <td></td> <td>20</td> <td>0</td> <td></td> <td></td> </tr> </tbody> </table>	Colonne	De Column	Créé	Type de BiO	Type	Modèle de date (Créé)	Longueur	Precision	Par défaut	Commentaire	CODE_CLIENT	CODE_CLIENT		<input type="checkbox"/> NUMBER	BigDecimal		20	0			MARCHE	MARCHE		<input type="checkbox"/> VARCHAR(40)	String		20	0			SECOMENT	SECOMENT		<input type="checkbox"/> VARCHAR(40)	String		20	0			PROFESSION	PROFESSION		<input type="checkbox"/> VARCHAR(40)	String		20	0			VILLECLIENT	VILLECLIENT		<input type="checkbox"/> VARCHAR(40)	String		20	0			SOUS_SECOMENT	SOUS_SECOMENT		<input type="checkbox"/> VARCHAR(40)	String		20	0			DATE_OUVERTURE	DATE_OUVERTURE		<input type="checkbox"/> DATE	Date		20	0			DATE_MODIFICATION	DATE_MODIFICATION		<input type="checkbox"/> DATE	Date		20	0			CODE_GEOGRAPHIQUE	CODE_GEOGRAPHIQUE		<input type="checkbox"/> NUMBER	BigDecimal	"dd/MM/yyyy"	20	0			ADRESSE_CLIENT	ADRESSE_CLIENT		<input type="checkbox"/> NUMBER	BigNumber	"dd/MM/yyyy"	20	0			DATE_NAISSANCE	DATE_NAISSANCE		<input type="checkbox"/> VARCHAR(40)	String		20	0		
Colonne	De Column	Créé	Type de BiO	Type	Modèle de date (Créé)	Longueur	Precision	Par défaut	Commentaire																																																																																																															
CODE_CLIENT	CODE_CLIENT		<input type="checkbox"/> NUMBER	BigDecimal		20	0																																																																																																																	
MARCHE	MARCHE		<input type="checkbox"/> VARCHAR(40)	String		20	0																																																																																																																	
SECOMENT	SECOMENT		<input type="checkbox"/> VARCHAR(40)	String		20	0																																																																																																																	
PROFESSION	PROFESSION		<input type="checkbox"/> VARCHAR(40)	String		20	0																																																																																																																	
VILLECLIENT	VILLECLIENT		<input type="checkbox"/> VARCHAR(40)	String		20	0																																																																																																																	
SOUS_SECOMENT	SOUS_SECOMENT		<input type="checkbox"/> VARCHAR(40)	String		20	0																																																																																																																	
DATE_OUVERTURE	DATE_OUVERTURE		<input type="checkbox"/> DATE	Date		20	0																																																																																																																	
DATE_MODIFICATION	DATE_MODIFICATION		<input type="checkbox"/> DATE	Date		20	0																																																																																																																	
CODE_GEOGRAPHIQUE	CODE_GEOGRAPHIQUE		<input type="checkbox"/> NUMBER	BigDecimal	"dd/MM/yyyy"	20	0																																																																																																																	
ADRESSE_CLIENT	ADRESSE_CLIENT		<input type="checkbox"/> NUMBER	BigNumber	"dd/MM/yyyy"	20	0																																																																																																																	
DATE_NAISSANCE	DATE_NAISSANCE		<input type="checkbox"/> VARCHAR(40)	String		20	0																																																																																																																	

Figure 61 : Structure de la table « Age »

2. La deuxième étape consiste à l'utilisation du premier composant **tMap** pour récupérer les dates de naissance sous la forme dd/mm/yyyy puis l'utilisation de composant **tFilterRow** pour filtrer les lignes avec une valeur « null ». Enfin l'utilisation du deuxième composant **tMap** pour récupérer les dates de naissance qui sont comprises entre 0 et 100 et l'utilisation du composant **tConvertType** pour convertir les dates de naissance au type string afin de les utilisées dans la récupération des anomalies. Les figures de 62 à 67 montrent ces composants.

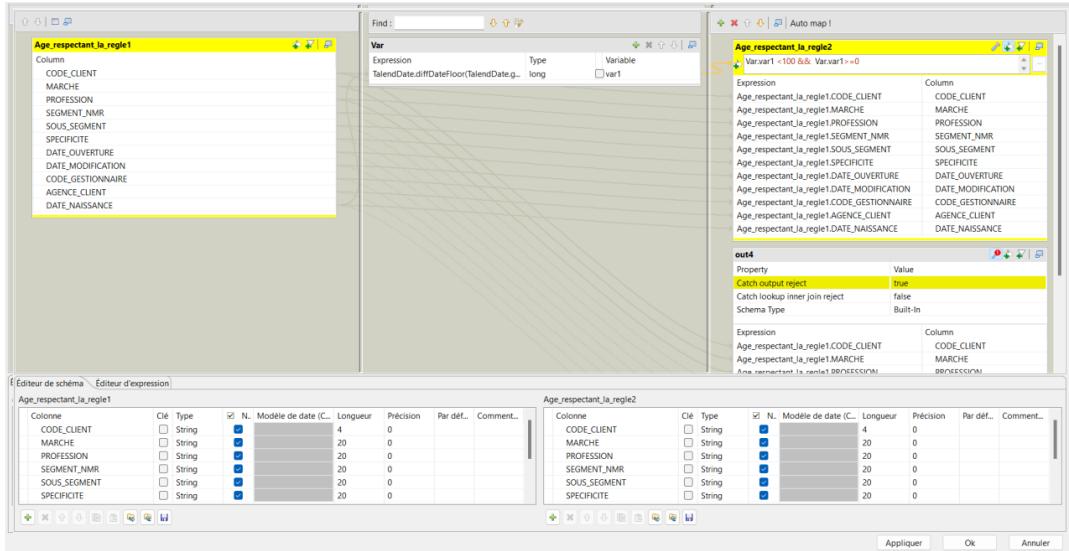


Figure 62 : Composant « **tMap1** »

```

Fonctions.isDateFormatValid
(Age.DATE_NAISSANCE, "EN")?
TalendDate.parseDateLocale("EEE MMM dd
HH:mm:ss zzz
yyyy", Age.DATE_NAISSANCE, "EN") :null

```

Var	Vale
Age.CODE_CLIENT	null
Age.MARCHE	null
Age.SEGMENT_NMR	null
Age.PROFESSION	null
Age.SOUS_SEGMENT	null
Age.SPECIFIQUE	null
Age.DATE_NAISANCE	null

Figure 63 : Expression pour la règle de gestion qui consiste à avoir les dates de naissance sous la forme dd/mm/yyyy

Conditions	Colonne d'entrée	Fonction	Opérateur	Valeur
	DATE_NAISANCE	Vider	Différent de	null

Figure 64 : Composant « **tFiltreRow** »

Figure 65 : Composant « tMap2 »

Figure 66 : Expression pour la règle de gestion qui consiste à avoir les dates de naissance dont l'âge est compris entre 0 et 100

Figure 67 : Composant « tConvertType »

3. La troisième étape consiste à charger les tables dans la base de données ORACLE. Pour se faire nous avons utilisé le composant **tOracleOutput**. La figure 68 montre la configuration du composant **tOracleOutput** pour la table « Age_Correct ».

Figure 68 : Composant **tOracleOutput** pour la table « Age_Correctes »

Figure 69 : Composant « tMap3 »

Ce composant permet de faire le mapping entre les deux tables « Age » et « Age_Correctes » afin de récupérer les lignes qui ne figurent pas dans les deux tables, ces lignes présentent des anomalies de date de naissance. Une seule sortie renvoyée par ce composant tMap, elle contient les lignes rejetées par la jointure entre les deux tables d'entrées. La figure 70 la configuration du composant tOracleOutput pour la table « Anomalies_Age ».

Figure 70 : Composant tOracleOutput pour la table « Anomalies_Age »

Les figures de 71 à 73 montrent l'exécution du Job « T_Age » (figure 71) ainsi que le chargement des deux tables « Age_correctes » et « Anomalies_Age » dans la base ORACLE (figure 72 et 73).

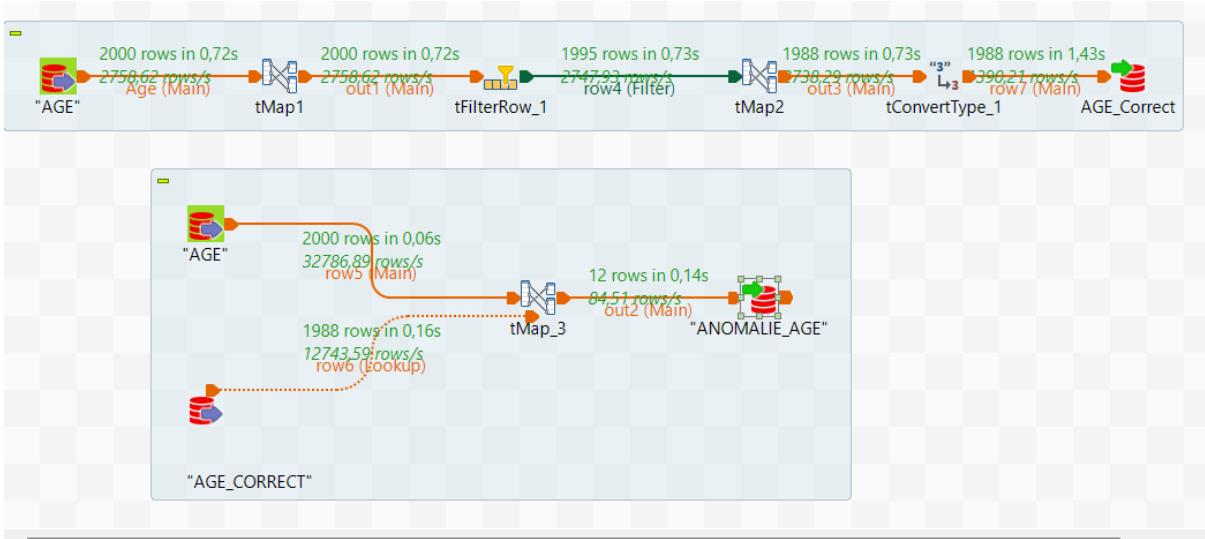


Figure 71 : Exécution du job « T_age »

	CODE_CLIENT	MARCHE	PROFESSION	SEGMENT_NMR	SOUS_SEGMENT	SPECIFICITE	DATE_OUVERTURE	DATE_MODIFICATION	CODE_GESTIONNAIRE	AGENCE_CLIENT	DATE_NAISANCE
1	1	TPE	F14	S5	14	TR GEI	10/06/15	20/07/15	612	16	15/07/96
2	10	Particuliers	F3	S1	13	DIGITAL TSF	27/03/10	06/05/10	989	151	20/01/95
3	100	Profession libérale	F9	S3	11	DIGITAL TSF	05/11/20	15/12/20	383	62	06/05/43
4	1000	Profession libérale	F9	S5	15	DIGITAL TSF	30/04/18	09/06/18	334	150	19/02/88
5	1002	TPE	F14	S1	12	OFF SHORE	17/06/16	27/07/16	561	186	30/01/75
6	1003	TPE	F6	S5	13	VIP	11/06/19	21/07/19	309	54	29/06/88
7	1004	TPE	F15	S5	13	BANQUE DIG...	26/10/21	05/12/21	248	53	25/04/59
8	1005	Entreprise	F16	S1	12	DIGITAL TSF	27/06/17	06/08/17	328	145	31/03/05
9	1006	Particuliers	F4	S5	13	TR GEI	25/11/22	04/01/23	973	223	19/10/82
10	1007	Profession libérale	F11	S2	11	BANQUE DIG...	04/02/08	15/03/08	918	96	25/07/04
11	1008	Profession libérale	F9	S1	11	VIP	09/01/20	18/02/20	866	20	02/05/75
12	1009	TPE	F7	S2	14	DIGITAL TSF	00/03/14	17/04/14	846	11	26/04/13
13	101	Entreprise	F17	S3	14	OFF SHORE	03/03/13	12/04/13	765	149	21/03/09
14	1010	Entreprise	F17	S5	13	OFF SHORE	29/01/15	10/03/15	910	116	23/03/28
15	1011	Entreprise	F16	S4	11	BANQUE DIG...	27/07/15	05/09/15	881	3	25/08/90
16	1012	Profession libérale	F11	S5	13	DIGITAL TSF	13/10/19	22/11/19	402	104	02/07/83
17	1013	Entreprise	F16	S3	13	TR GEI	23/02/10	04/04/10	519	186	23/12/71
18	1014	TPE	F14	S5	15	BANQUE PRIVEE	10/01/11	19/02/11	180	154	30/05/09
19	1015	TPE	F8	S4	12	TR GEI	18/11/14	28/12/14	730	54	29/11/98
20	1016	TPE	F13	S5	15	BANQUE PRIVEE	20/10/11	07/12/11	165	162	26/10/28
21	1017	Entreprise	F16	S3	12	BANQUE DIG...	01/09/15	11/10/15	463	12	07/02/05
22	1018	Particuliers	F4	S3	15	DIGITAL TSF	05/01/08	14/02/08	193	154	07/01/47
23	1019	Profession libérale	F10	S5	12	BANQUE PRIVEE	30/11/18	09/01/19	200	44	23/01/89
24	102	Particuliers	F3	S3	15	VIP	19/12/22	28/01/23	638	115	06/10/80
25	1020	Profession libérale	F10	S1	12	TR GEI	02/06/22	12/07/22	844	121	02/07/47
26	1021	TPE	F8	S5	15	TR GEI	18/06/14	28/07/14	259	65	16/03/60
27	1022	Profession libérale	F11	S1	14	OFF SHORE	30/10/15	09/12/15	305	171	01/09/97
28	1023	TPE	F17	S1	14	BANQUE DIG...	11/04/16	31/05/16	451	??	21/01/75

Figure 72 : Chargement de la table « Age_correctes »

The screenshot shows the Oracle SQL Developer interface. On the left, the 'Connexions' sidebar lists 'ANOMALIE_AGE' and 'Reports'. The main area displays a table titled 'ANOMALIE_AGE' with 13 rows of data. The columns are: #ID, PROFESSION, SEGMENT_NMR, SOUS_SEGMENT, SPECIFIQUE, DATE_OUVERTURE, DATE_MODIFICATION, CODE_GESTIONNAIRE, AGENCE_CLIENT, and DATE_NAISANCE. The data includes various professions like 'session liberale', 'iculiers', 'prise', etc., and segments like 'VIP', 'BANQUE PRIVEE', 'DIGITAL TSF', etc.

#ID	PROFESSION	SEGMENT_NMR	SOUS_SEGMENT	SPECIFIQUE	DATE_OUVERTURE	DATE_MODIFICATION	CODE_GESTIONNAIRE	AGENCE_CLIENT	DATE_NAISANCE	
1	session liberale	P11	S3	I5	TR GEI	28/12/11	06/02/12	459	221	Fri Aug 17 00:00:00 WAT 1900
2	iculiers	P2	S1	I4	VIP	02/10/17	11/11/17	611	30	mo9/09/1966
3	prise	P17	S5	I3	TR GEI	05/05/12	14/06/12	977	66	*21/12/1958
4		P13	S1	I3	BANQUE PRIVEE	16/07/09	08/04/23	714	200	20/01/2007
5		P12	S1	I1	DIGITAL TSF	06/12/13	09/04/23	845	43	15/78/1939
6		P12	S4	I3	TR GEI	18/06/17	28/07/17	315	105	25/08/1962
7	session liberale	P10	S5	I4	OFF SHORE	28/02/18	09/04/18	571	231	Tue Apr 17 00:00:00 WAT 1923
8	session liberale	P10	S2	I4	OFF SHORE	28/02/18	09/04/18	595	137	Sun Mar 11 00:00:00 WAT 1923
9	session liberale	P11	S4	I2	TR GEI	30/10/16	09/12/16	243	106	Sat Mar 10 00:00:00 WAT 1923
10	iculiers	P2	S5	I1	BANQUE DIGITALE	14/06/19	24/07/19	330	200	Sun May 06 00:00:00 WAT 1923
11	session liberale	P9	S3	I1	DIGITAL TSF	18/08/21	27/09/21	439	186	Sat Mar 03 00:00:00 WAT 1923
12		P6	S3	I2	OFF SHORE	15/01/09	24/02/09	366	218	Sat Feb 10 00:00:00 WAT 1900
13	iculiers	P2	S5	I3	OFF SHORE	06/04/12	16/05/12	815	136	Thu Oct 16 00:00:00 WAT 1902

Figure 73 : Chargement de la table « Anomalies_Age »

2.3.3 Application des règles de gestion pour transformer les données liées aux numéros de CIN

La figure 74 montre l'environnement du job concernant les transformations effectuées sur les données importées de la table « CIN » qui a été chargée dans l'étape précédente.

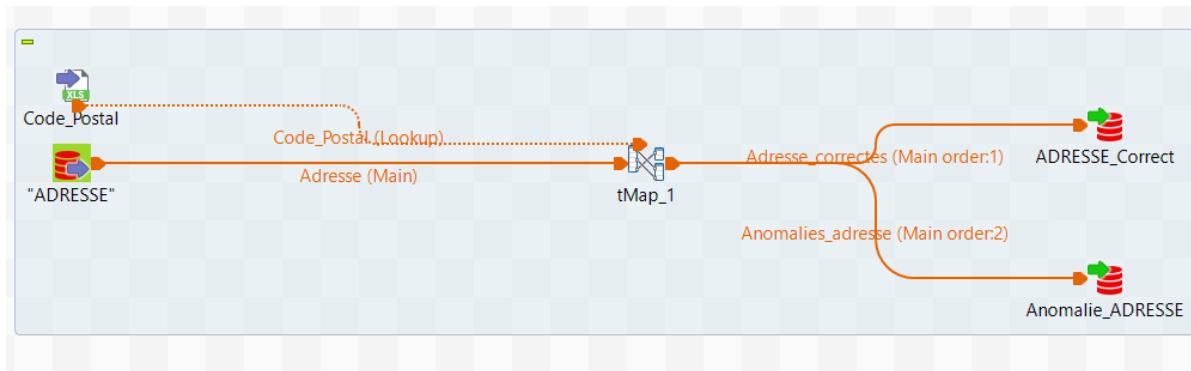


Figure 74 : Environnement de travail du job « T_CIN »

Ce job permet d'appliquer les règles de gestion pour les numéros de CIN des clients. Un numéro de CIN est considéré comme numéro correcte s'il:

- Contient exactement 8 chiffres
- Commence par 0 ou 1
- Contient uniquement des chiffres

1. La première étape pour la réalisation de ce job consiste à importer la table CIN en utilisant le composant d'entrée **tOracleInput**. Les figures 75 et 76 montrent ce composant ainsi que son schéma.



Figure 75 : Table « CIN »

Schéma de "CIN"												
	Colonne utilisée	Colonne	Db Column	Clé	Type	Type ...	✓ N..	Modèle de ...	Longu...	Précis...	Par ...	Com...
<input checked="" type="checkbox"/>	CODE_CLIENT	CODE_CLIENT	CODE_CLIENT	<input type="checkbox"/>	String	VARC...	<input checked="" type="checkbox"/>		4	0		
<input checked="" type="checkbox"/>	MARCHE	MARCHE	MARCHE	<input type="checkbox"/>	String	VARC...	<input checked="" type="checkbox"/>		20	0		
<input checked="" type="checkbox"/>	PROFESSION	PROFESSION	PROFESSION	<input type="checkbox"/>	String	VARC...	<input checked="" type="checkbox"/>		20	0		
<input checked="" type="checkbox"/>	SEGMENT_NMR	SEGMENT_NMR	SEGMENT_NMR	<input type="checkbox"/>	String	VARC...	<input checked="" type="checkbox"/>		20	0		
<input checked="" type="checkbox"/>	SOUS_SEGMENT	SOUS_SEGMENT	SOUS_SEGMENT	<input type="checkbox"/>	String	VARC...	<input checked="" type="checkbox"/>		20	0		
<input checked="" type="checkbox"/>	SPECIFICITE	SPECIFICITE	SPECIFICITE	<input type="checkbox"/>	String	VARC...	<input checked="" type="checkbox"/>		20	0		
<input checked="" type="checkbox"/>	DATE_OUVERTURE	DATE_OUVERTURE	DATE_OUVERTURE	<input type="checkbox"/>	Date	DATE	<input checked="" type="checkbox"/>	"dd/MM/yy...	7	0		
<input checked="" type="checkbox"/>	DATE_MODIFICA...	DATE_MODIFICATI...	DATE_MODIFICATI...	<input type="checkbox"/>	Date	DATE	<input checked="" type="checkbox"/>	"dd/MM/yy...	7	0		
<input checked="" type="checkbox"/>	CODE_GESTION...	CODE_GESTIONNAI...	CODE_GESTIONNAI...	<input type="checkbox"/>	String	VARC...	<input checked="" type="checkbox"/>		3	0		
<input checked="" type="checkbox"/>	AGENCE_CLIENT	AGENCE_CLIENT	AGENCE_CLIENT	<input type="checkbox"/>	String	VARC...	<input checked="" type="checkbox"/>		4	0		
<input checked="" type="checkbox"/>	NUM_CIN	NUM_CIN	NUM_CIN	<input type="checkbox"/>	String	VARC...	<input checked="" type="checkbox"/>		10	0		

Figure 76 : Structure de la table « CIN »

2. La deuxième étape consiste à l'utilisation du composant **tMap** pour appliquer les règles de gestion afin d'avoir deux sorties : une table pour les informations qui contiennent des anomalies pour les numéros de CIN et une autre pour les numéros de CIN correctes. La figure 77 montre le composant **tMap**.

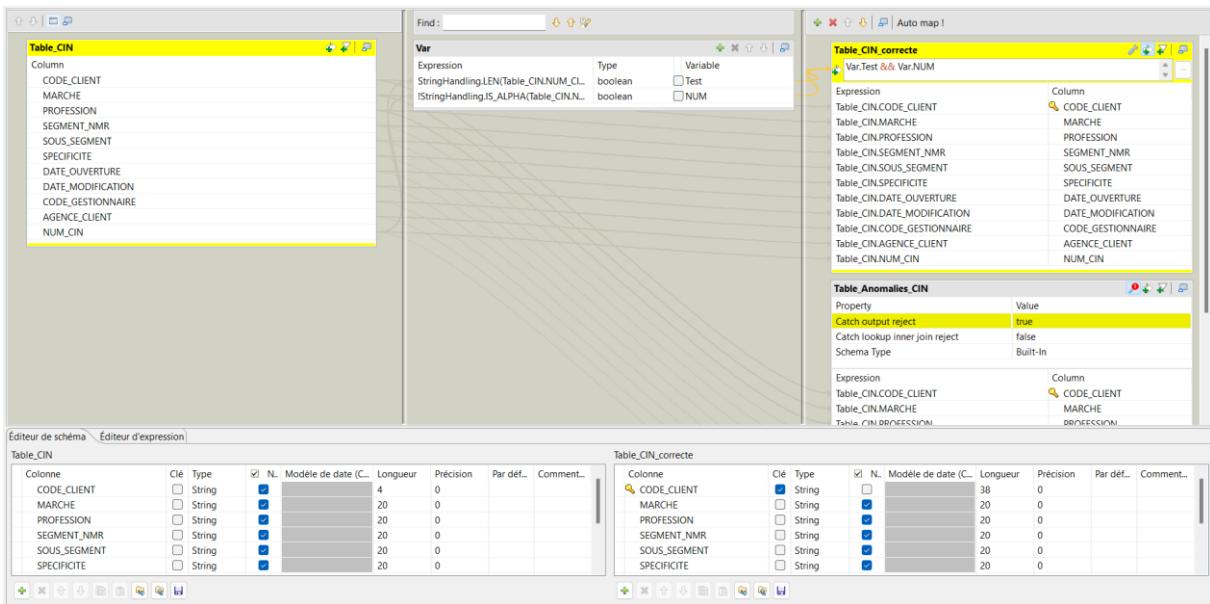


Figure 77 : Composant tMap pour la transformation des numéros de CIN

La figure 77 présente l'interface de l'éditeur de mapping du composant tMap. À gauche, nous pouvons voir le schéma de la table d'entrée nommées « Age ». À droite, le schéma des tables de sortie « CIN_correct » et « Anomalies_CIN ». Pour extraire les données de cette table d'entrée, il est nécessaire de faire le mapping entre la table d'entrée et les deux tables de sorties. Pour ce faire, il suffit de faire glisser les colonnes de la table « CIN » vers les deux tables de sorties et d'activer l'option ‘catch output reject’ sur ‘True’ pour récupérer les numéros de CIN qui ne respectent pas les règles de gestion qui sont présent dans les figures 78 et 79.

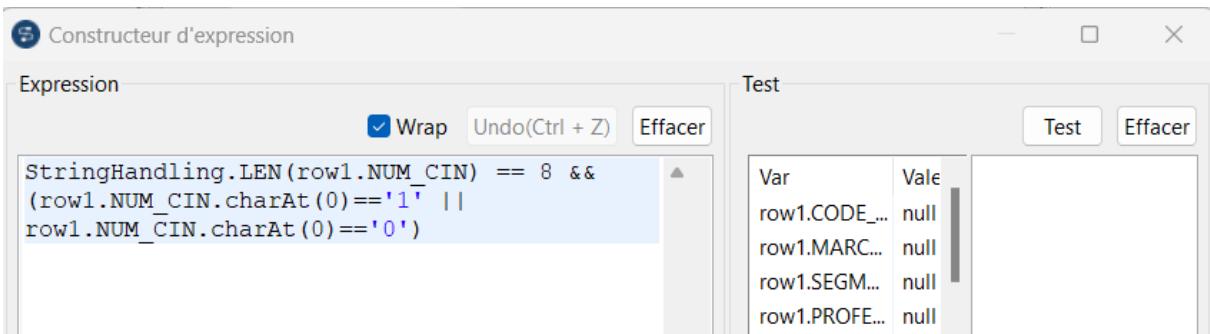


Figure 78 : Expression pour la règle de gestion qui consiste à avoir les numéros de CIN qui commence par 0 et 1 et leur longueur est de 8 chiffres

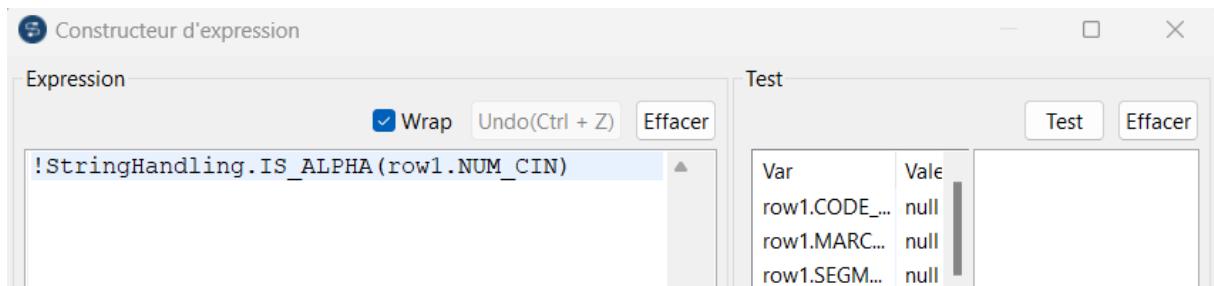


Figure 79 : Expression pour la règle de gestion qui consiste à avoir les numéros de CIN qui contiennent que des numéros

3. La troisième étape consiste à charger les tables dans la base de données ORACLE. Pour se faire nous avons utilisé le composant **tOracleOutput**. Les figures 80 et 81 montrent la configuration des deux composants **tOracleOutput** pour chaque table.

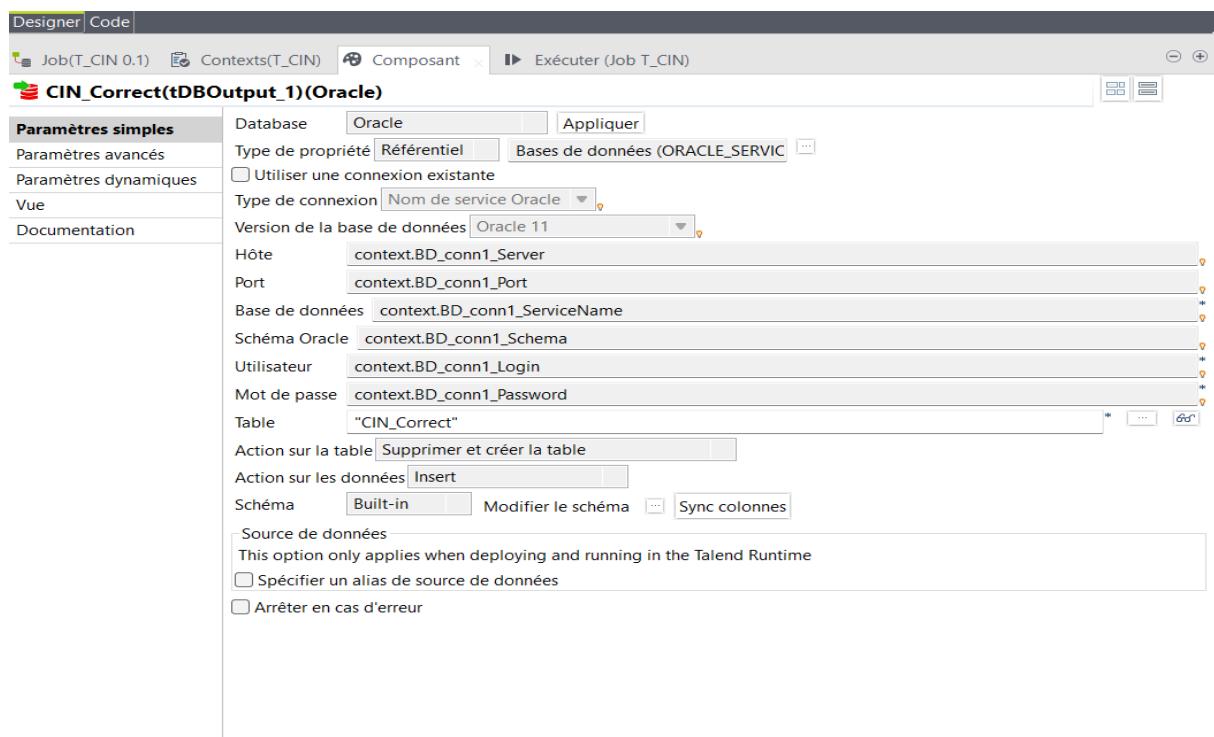


Figure 80 : Composant **tOracleOutput** pour la table « CIN_Correct »

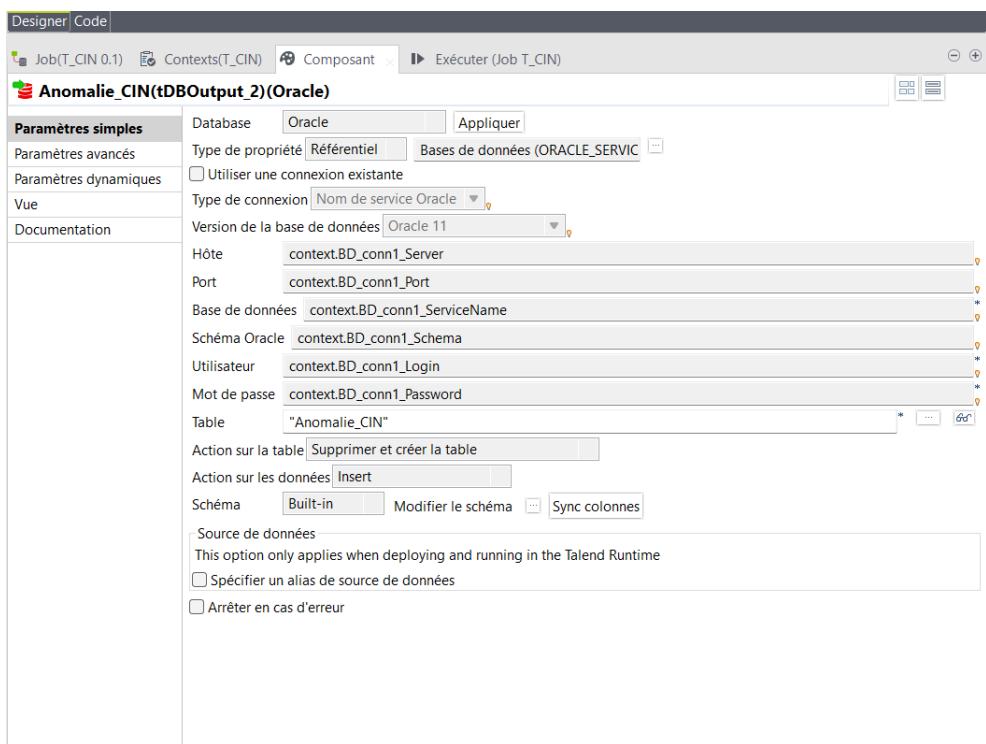


Figure 81 : Composant *tOracleOutput* pour la table « Anomalies_CIN »

Les figures de 82 à 84 montrent l'exécution du Job « T_CIN » (figure 82) ainsi que le chargement des deux tables « CIN_correct » et « Anomalies_CIN » dans la base ORACLE (figure 83 et 84).

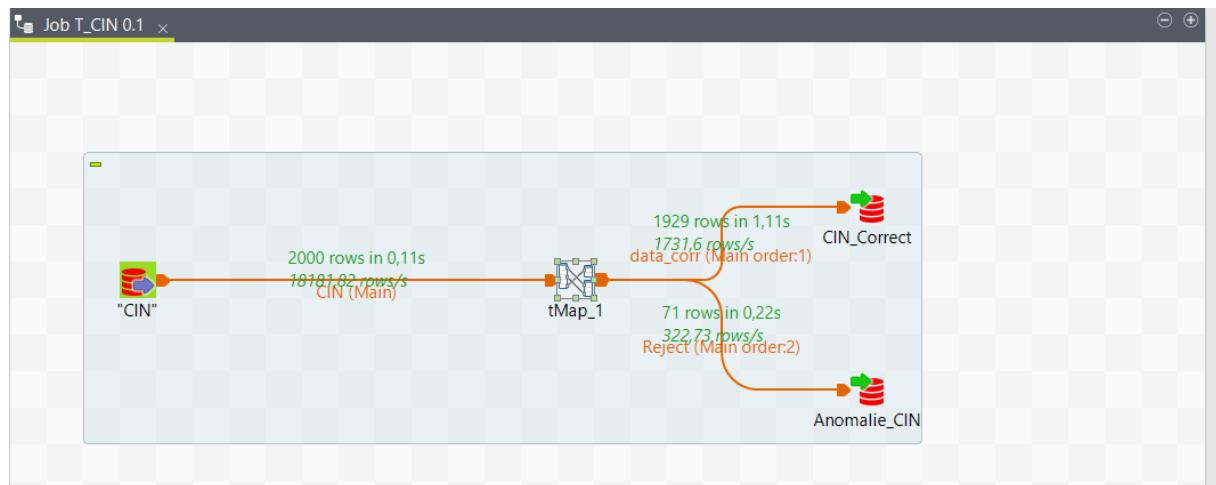


Figure 82 : Exécution du job « T_CIN »

The screenshot shows the Oracle SQL Developer interface with the 'script_MAJ_Mensuelle.sql' script running. The 'CIN_CORRECT' table is displayed in the main pane, showing 28 rows of data. The columns are: CODE_CLIENT, MARCHE, PROFESSION, SEGMENT_NMR, SOUS_SEGMENT, SPECIFICITE, DATE_OUVERTURE, DATE_MODIFICATION, CODE_GESTIONNAIRE, AGENCE_CLIENT, and NUM_CIN. The data includes various client codes, market segments like TPE, TRE, and Particuliers, and specific codes like P1, P2, P3, etc.

Figure 83 : Chargement de la table « CIN_correctes »

The screenshot shows the Oracle SQL Developer interface with the 'script_MAJ_Mensuelle.sql' script running. The 'ANOMALIE_CIN' table is displayed in the main pane, showing 43 rows of data. The columns are: CODE_CLIENT, MARCHE, PROFESSION, SEGMENT_NMR, SOUS_SEGMENT, SPECIFICITE, DATE_OUVERTURE, DATE_MODIFICATION, CODE_GESTIONNAIRE, AGENCE_CLIENT, and NUM_CIN. The data includes various client codes, market segments like TPE, TRE, and Particuliers, and specific codes like P1, P2, P3, etc., along with anomaly codes like VIP, OFF SHORE, and TR GEI.

Figure 84 : Chargement de la table « Anomalies_CIN »

2.3.4 Application des règles de gestion pour transformer les données liées aux noms et prénoms des clients

La figure 85 montre l'environnement du job concernant les transformations effectuées sur les données importées de la table « Nom_Prenom » qui a été chargée dans l'étape précédente.

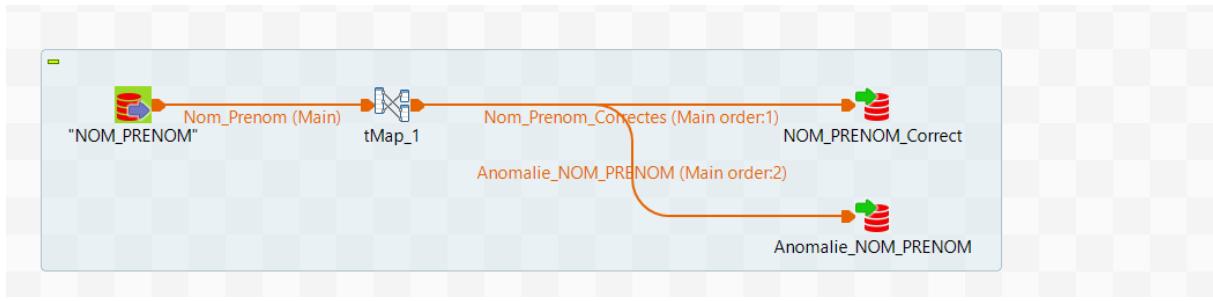


Figure 85 : Environnement de travail du job « T_Nom_Prenom »

Ce job permet d'appliquer les règles de gestion pour les numéros de CIN des clients. Un nom et un prénom sont considérés comme nom et prénom corrects s'ils :

-Ne sont pas null (si l'un des deux ne sont pas null il n'y a pas d'anomalie)

1. La première étape pour la réalisation de ce job consiste à importer la table « Nom_Prenom » en utilisant le composant d'entrée **tOracleInput**. Les figures 86 et 87 montrent ce composant ainsi que son schéma :



Figure 86 : Table « Nom_Prenom »

Schéma de "NOM_PRENOM"

"NOM_PRENOM"

Colonne utilisée	Colonne	Db Column	Clé	Type	Type ...	N.	Modèle de...	Long...	Précis...	Par ...	Com...
<input checked="" type="checkbox"/>	CODE_CLIENT	CODE_CLIENT	<input checked="" type="checkbox"/>	BigD...	NUM...	<input type="checkbox"/>		38	0		
<input checked="" type="checkbox"/>	MARCHE	MARCHE	<input type="checkbox"/>	String	VARC...	<input checked="" type="checkbox"/>		20	0		
<input checked="" type="checkbox"/>	SEGMENT	SEGMENT	<input type="checkbox"/>	String	VARC...	<input checked="" type="checkbox"/>		20	0		
<input checked="" type="checkbox"/>	PROFESSION	PROFESSION	<input type="checkbox"/>	String	VARC...	<input checked="" type="checkbox"/>		20	0		
<input checked="" type="checkbox"/>	SEGMENT_NMR	SEGMENT_NMR	<input type="checkbox"/>	String	VARC...	<input checked="" type="checkbox"/>		20	0		
<input checked="" type="checkbox"/>	SOUS_SEGMENT	SOUS_SEGMENT	<input type="checkbox"/>	String	VARC...	<input checked="" type="checkbox"/>		20	0		
<input checked="" type="checkbox"/>	DATE_OUVERTURE	DATE_OUVERTURE	<input type="checkbox"/>	Date	DATE	<input checked="" type="checkbox"/>	"dd/MM/y...	20	0		
<input checked="" type="checkbox"/>	DATE_MODIFICATION	DATE_MODIFICATION	<input type="checkbox"/>	Date	DATE	<input checked="" type="checkbox"/>	"dd/MM/y...	20	0		
<input checked="" type="checkbox"/>	CODE_GESTIONNAIRE	CODE_GESTIONNAIRE	<input type="checkbox"/>	BigD...	NUM...	<input checked="" type="checkbox"/>		38	0		
<input checked="" type="checkbox"/>	AGENCE_CLIENT	AGENCE_CLIENT	<input type="checkbox"/>	BigD...	NUM...	<input checked="" type="checkbox"/>		38	0		
<input checked="" type="checkbox"/>	NOM	NOM	<input type="checkbox"/>	String	VARC...	<input checked="" type="checkbox"/>		10	0		
<input checked="" type="checkbox"/>	PRENOM	PRENOM	<input type="checkbox"/>	String	VARC...	<input checked="" type="checkbox"/>		10	0		

Figure 87 : Structure de la table « Nom_Prenom »

2. La deuxième étape consiste à l'utilisation du composant **tMap** pour récupérer les noms et prénoms corrects ainsi que ceux présentant des anomalies. La figure 88 montre ce composant.

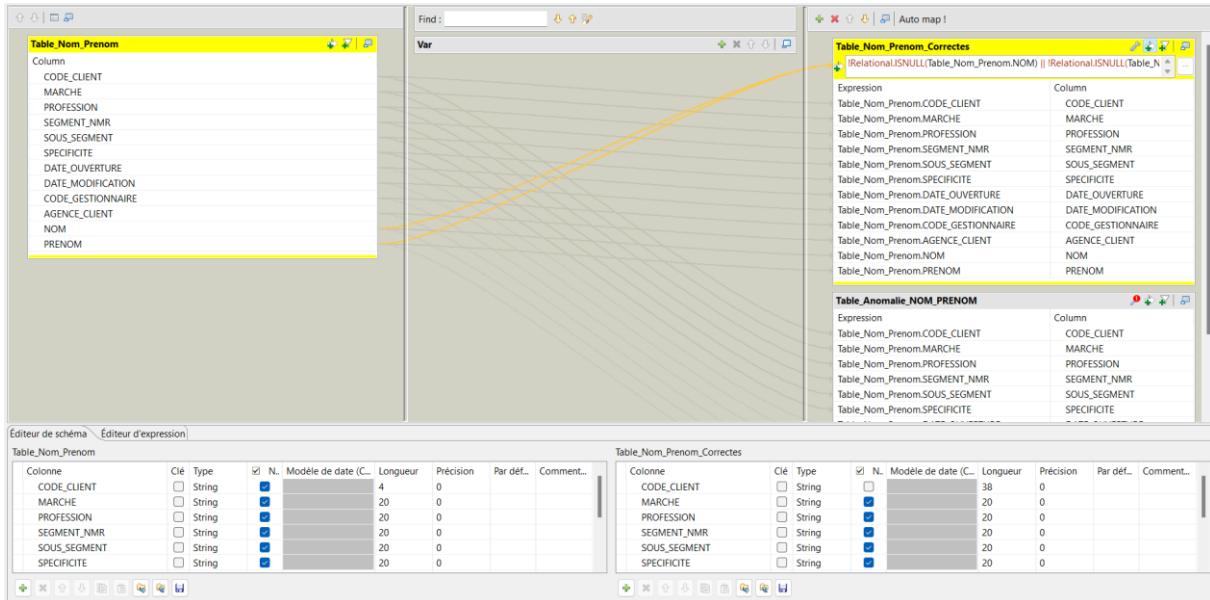


Figure 88 : Composant « **tMap_Nom_Prenom** »

La figure 88 présente l'interface de l'éditeur de mapping du composant **tMap**. À gauche, nous pouvons voir le schéma de la table d'entrée nommées « **Nom_Prenom** ». À droite, le schéma des tables de sortie « **Nom_Prenom_corrects** » et « **Anomalies_Nom_Prenom** ». Pour extraire les données de cette table d'entrée, il est nécessaire de faire le mapping entre la table d'entrée et les deux tables de sorties. Pour ce faire, il suffit de faire glisser les colonnes de la table « **Nom_Prenom** » vers les deux tables de sorties et d'activer l'option ‘catch output reject’ sur ‘True’ pour récupérer les noms et prénom qui ne respectent pas la règle de gestion qui est présente dans la figure 89.

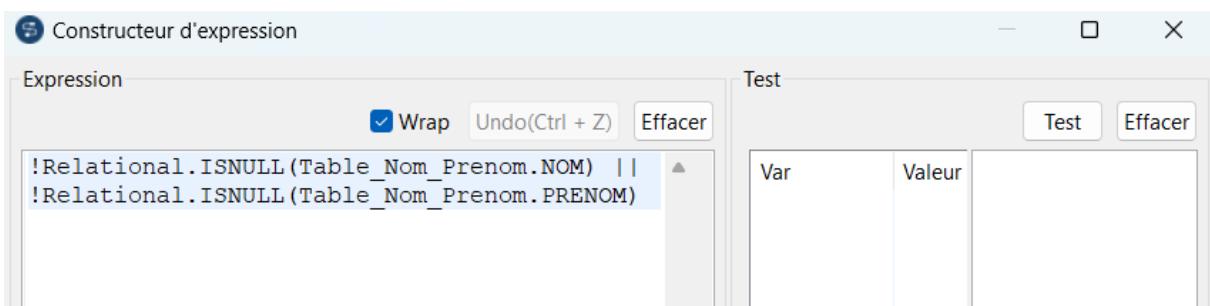


Figure 89 : Expression pour la règle de gestion qui consiste à avoir les noms et prénoms non null

3. La troisième étape consiste à charger les tables dans la base de données ORACLE. Pour se faire nous avons utilisé le composant **tOracleOutput**. Les figures 90 et 91 montrent la configuration des deux composants **tOracleOutput** pour chaque table.

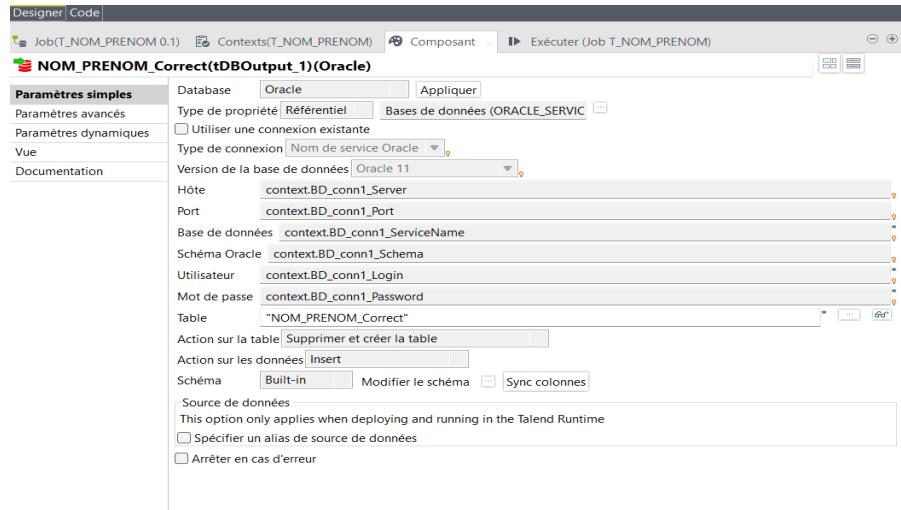


Figure 90 : Composant **tOracleOutput** pour la table « Nom_Prenom_Corrects »

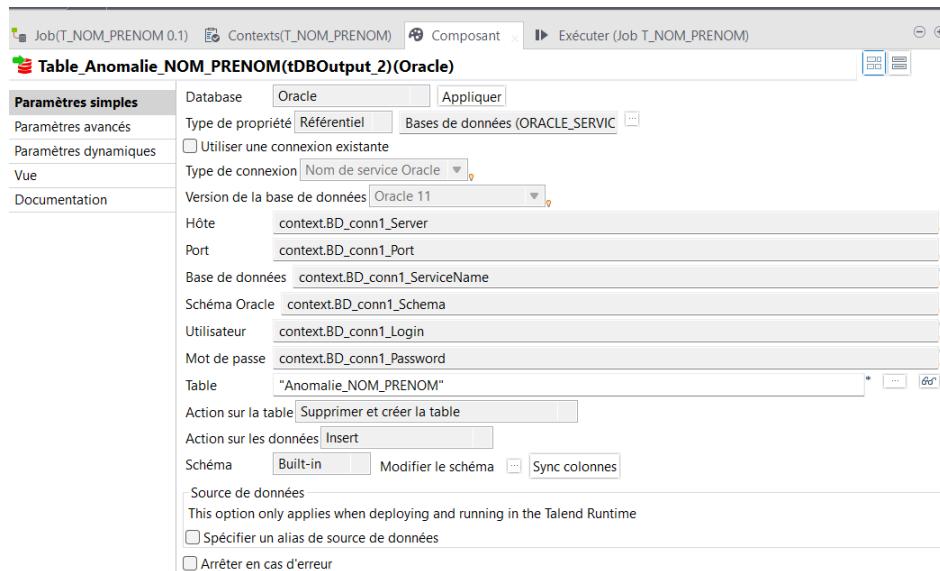


Figure 91 : Composant **tOracleOutput** pour la table « Anomalies_Nom_Prenom »

Les figures de 92 à 94 montrent l'exécution du Job « T_CIN » (figure 92) ainsi que le chargement des deux tables « Nom_Prenom_Corrects » et "Anomalies_Nom_Prenom" dans la base ORACLE (figure 93 et 94) :

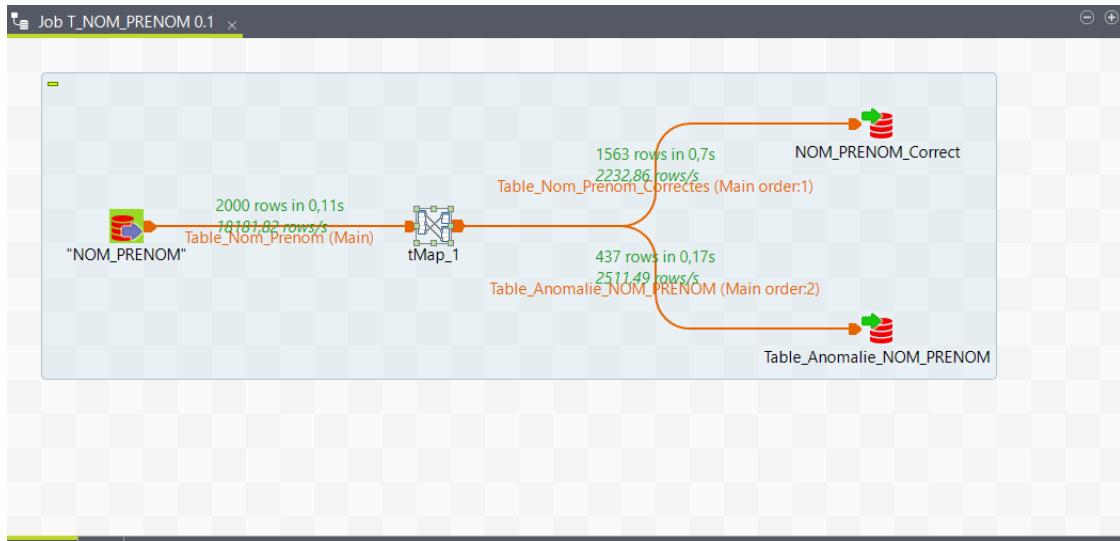


Figure 92 : Exécution du job « T_Nom_Prenom »

IDENT	MARCHE	PROFESSION	SEGMENT_NMR	SOUS_SEGMENT	SPECIFIQUE	DATE_OUVERTURE	DATE_MODIFICATION	CODE_GESTIONNAIRE	AGENCE_CLIENT	NOM	PRENOM
1	Entreprise	P16	55	13	BANQUE DIGITALE	20/03/14	29/04/14	750	137	(null)	Oms
2	TPE	P13	52	14	TR GEI	11/03/08	20/04/08	958	141	Shabri	Meidi
3	TPE	P13	54	14	TR GEI	05/03/13	14/04/13	921	74	(null)	Anis
4	TRE	P7	54	11	BANQUE PRIVEE	24/07/16	02/09/16	306	135	(null)	Sami
5	Profession libérale	P9	53	12	TR GEI	26/05/19	05/07/19	843	104	(null)	Nadhel
6	TPE	P12	55	14	VIP	20/12/14	29/01/15	944	187	(null)	Rahma
7	Entreprise	P16	51	12	DIGITAL TSF	03/07/17	12/08/17	883	65	(null)	Meidi
8	Entreprise	P16	52	13	BANQUE PRIVEE	08/06/22	18/07/22	483	134	Maaji	(null)
9	Profession libérale	P10	51	15	BANQUE DIGITALE	17/02/10	29/03/10	294	161	(null)	Nadhem
10	TRE	P7	55	11	OFF SHORE	19/07/20	28/08/20	963	232	Otmane	(null)
11	Entreprise	P16	52	15	DIGITAL TSF	04/11/17	14/12/17	724	182	(null)	Hanen
12	TPE	P14	55	13	TR GEI	26/06/16	05/10/16	179	147	Maari	Mehdi
13	TPE	P15	52	14	VIP	06/07/16	15/08/16	403	159	Cherni	Abdelhamid
14	TPE	P14	55	12	VIP	01/05/15	10/06/15	315	94	(null)	Maram
15	Profession libérale	P9	54	15	BANQUE DIGITALE	15/03/18	24/04/18	256	44	(null)	Loffi
16	Profession libérale	P11	54	14	DIGITAL TSF	01/12/22	10/01/23	454	165	Zribi	(null)
17	TRE	P8	55	11	BANQUE PRIVEE	31/10/08	10/12/08	837	125	Zayed	Amar
18	TRE	P7	52	11	OFF SHORE	14/07/08	23/08/08	919	5	Berriche	Yosra
19	Profession libérale	P11	54	11	OFF SHORE	28/06/19	07/08/19	514	79	(null)	Khalil
20	TRE	P6	54	12	OFF SHORE	26/03/13	05/05/13	356	235	Toumi	Haissa
21	TPE	P14	55	12	OFF SHORE	21/11/08	31/12/08	981	161	Hammani	Kais
22	Profession libérale	P11	55	15	TR GEI	31/03/10	30/04/10	222	187	(null)	Meidi
23	Profession libérale	P10	53	13	BANQUE PRIVEE	07/08/15	16/09/15	454	12	Mabrouk	(null)
24	Particuliers	P4	81	13	BANQUE DIGITALE	05/01/23	14/02/23	580	84	Bouahmed	(null)
25	Profession libérale	P9	52	11	BANQUE PRIVEE	16/07/17	25/08/17	446	61	(null)	Alli
26	TRE	P6	53	14	BANQUE PRIVEE	25/05/08	04/07/08	938	69	Amri	Zied
27	Particuliers	P3	52	14	BANQUE DIGITALE	28/10/15	07/12/15	859	36	Shabri	Mehdi
28	yes	xx	xx	xx	xx	xx/xx/xx	xx/xx/xx	xx	xx	xx	xx

Figure 93 : Chargement de la table « Nom_Prenom_corrects »

IDENT	CODE_CLIENT	MARCHE	PROFESSION	SEGMENT_NMR	SOUS_SEGMENT	SPECIFIQUE	DATE_OUVERTURE	DATE_MODIFICATION	CODE_GESTIONNAIRE	AGENCE_CLIENT	NOM	PRENOM
1	878	Particuliers	P4	S2	14	BANQUE DIGITALE	02/01/23	11/02/23	656	27	(null)	(null)
2	880	Profession libérale	P10	S5	12	BANQUE DIGITALE	09/05/17	18/06/17	275	142	(null)	(null)
3	881	Entreprise	P16	S5	14	BANQUE DIGITALE	21/12/14	30/01/15	150	49	(null)	(null)
4	882	Profession libérale	P11	S3	12	BANQUE DIGITALE	03/12/16	12/01/17	402	69	(null)	(null)
5	895	TPE	P13	S5	11	BANQUE DIGITALE	23/07/08	01/09/08	447	49	(null)	(null)
6	897	Particuliers	P1	S1	13	DIGITAL TSF	24/11/12	03/01/13	729	179	(null)	(null)
7	898	TRE	P6	S4	14	BANQUE PRIVEE	13/02/20	24/03/20	976	28	(null)	(null)
8	899	TRE	P7	S3	15	TR GEI	24/04/20	03/06/20	908	115	(null)	(null)
9	901	TRE	P7	S4	12	BANQUE DIGITALE	30/04/18	09/06/18	526	28	(null)	(null)
10	912	Particuliers	P4	S2	14	DIGITAL TSF	14/02/13	26/03/13	204	56	(null)	(null)
11	913	TPE	P15	S1	13	BANQUE DIGITALE	29/04/21	08/06/21	677	138	(null)	(null)
12	921	TRE	P6	S4	14	VIP	24/06/14	03/08/14	598	36	(null)	(null)
13	923	TPE	P14	S5	14	BANQUE DIGITALE	29/09/20	08/11/20	270	133	(null)	(null)
14	931	Particuliers	P2	S4	11	DIGITAL TSF	12/11/08	22/12/08	371	65	(null)	(null)
15	933	TRE	P8	S5	15	DIGITAL TSF	17/11/22	27/12/22	635	40	(null)	(null)
16	935	Entreprise	P16	S2	15	BANQUE PRIVEE	06/07/14	15/08/14	764	135	(null)	(null)
17	940	Particuliers	P4	S1	13	BANQUE PRIVEE	11/01/18	20/02/18	785	228	(null)	(null)
18	953	TPE	P12	S5	11	TR GEI	06/05/11	15/06/11	120	33	(null)	(null)
19	956	Profession libérale	P10	S4	15	DIGITAL TSF	26/08/15	05/10/15	363	127	(null)	(null)
20	961	Profession libérale	P11	S4	14	TR GEI	20/02/15	01/04/15	501	37	(null)	(null)
21	964	Profession libérale	P9	S1	14	BANQUE DIGITALE	05/12/11	14/01/12	713	115	(null)	(null)
22	966	Entreprise	P17	S5	13	BANQUE DIGITALE	24/05/08	03/07/08	190	132	(null)	(null)
23	970	TPE	P15	S1	14	VIP	08/04/12	18/05/12	668	218	(null)	(null)
24	971	Profession libérale	P9	S4	14	VIP	30/03/09	09/05/09	789	176	(null)	(null)
25	975	Profession libérale	P9	S3	15	BANQUE PRIVEE	09/11/12	19/12/12	643	217	(null)	(null)
26	976	Profession libérale	P9	S5	11	DIGITAL TSF	21/03/09	30/04/09	389	30	(null)	(null)
27	981	TPE	P12	S2	12	VIP	27/01/11	08/03/11	543	49	(null)	(null)
28	982	Particuliers	P16	S1	14	DIGITAL TSF	31/12/16	06/03/17	829	153	(null)	(null)

Figure 94 : Chargement de la table « Anomalies_Nom_Prenom »

2.3.5 Application des règles de gestion pour transformer les données liées aux numéros de téléphone des clients

La figure 95 montre l'environnement du job concernant les transformations effectuées sur les données importées de la table « Tel » qui a été chargée dans l'étape précédente.

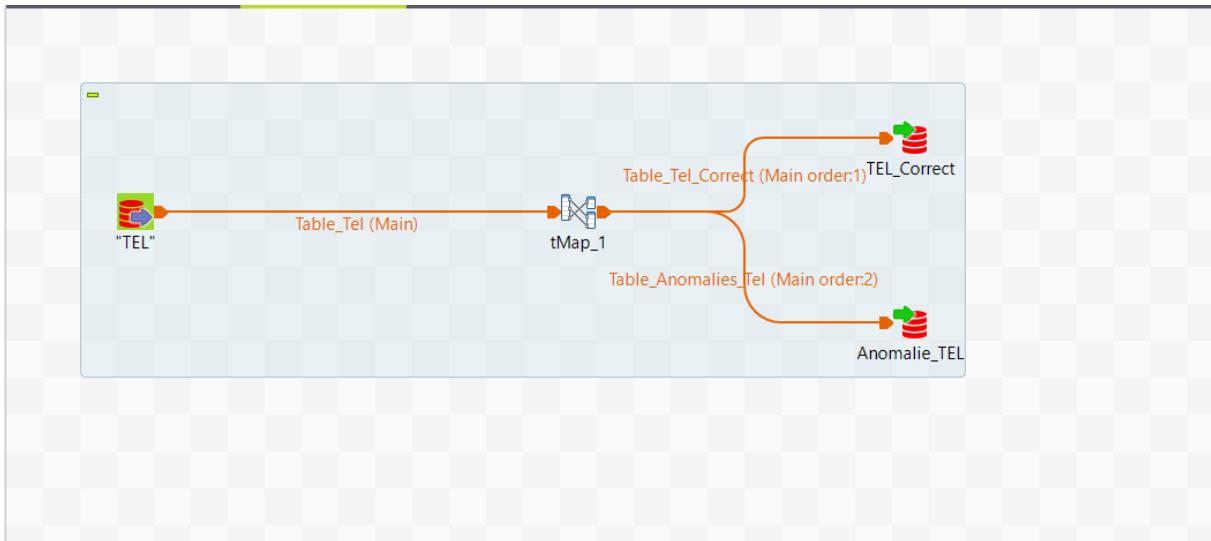


Figure 95 : Environnement de travail du job « T_Tel »

Ce job permet d'appliquer les règles de gestion pour les numéros de téléphone des clients. Un numéro de téléphone est correct si:

- Il contient exactement 8 chiffres
- Il est numérique (ne contient que des chiffres)
- Il ne commence pas par 7 ou 0 ou 3 ou 4

Ces règles vont être appliquées sur les numéros de téléphone fiables et portables. Si l'un des deux respecte la règle il n'y a pas d'anomalies.

1. La première étape pour la réalisation de ce job consiste à importer la table Tel en utilisant le composant d'entrée **tOracleInput**. Les figures 96 et 97 montrent ce composant ainsi que son schéma :



Figure 96 : Table « Tel »

Schéma de "TEL"

"TEL"

Colonne utilisée	Colonne	Db Column	Cle	Type	Type d...	N.	Modèle de d...	Longu...	Précisi...	Par d...	Comm...
✓	CODE_CLIENT	CODE_CLIENT	✓	BigDe...	NUMB...	✓		38	0		
✓	MARCHE	MARCHE		String	VARC...	✓		20	0		
✓	SEGMENT	SEGMENT		String	VARC...	✓		20	0		
✓	PROFESSION	PROFESSION		String	VARC...	✓		20	0		
✓	SEGMENT_NMR	SEGMENT_NMR		String	VARC...	✓		20	0		
✓	SOUS_SEGMENT	SOUS_SEGMENT		String	VARC...	✓		20	0		
✓	DATE_OUVERTURE	DATE_OUVERTURE		Date	DATE	✓	"dd/MM/yyyy"	20	0		
✓	DATE_MODIFICATION	DATE_MODIFICATION		Date	DATE	✓	"dd/MM/yyyy"	20	0		
✓	CODE_GESTIONNAIRE	CODE_GESTIONNAIRE		BigDe...	NUMB...	✓		38	0		
✓	AGENCE_CLIENT	AGENCE_CLIENT		BigDe...	NUMB...	✓		38	0		
✓	TELEPHONE_FIABLE	TELEPHONE_FIABLE		String	VARC...	✓		30	0		
✓	TELEPHONE_PORTABLE	TELEPHONE_PORTABLE		String	VARC...	✓		30	0		

OK Cancel

Figure 97 : Structure de la table « Tel »

2. La deuxième étape consiste à l'utilisation du composant tMap pour récupérer les numéros de téléphones corrects ainsi que ceux présentant des anomalies. La figure 98 montre ce composant.

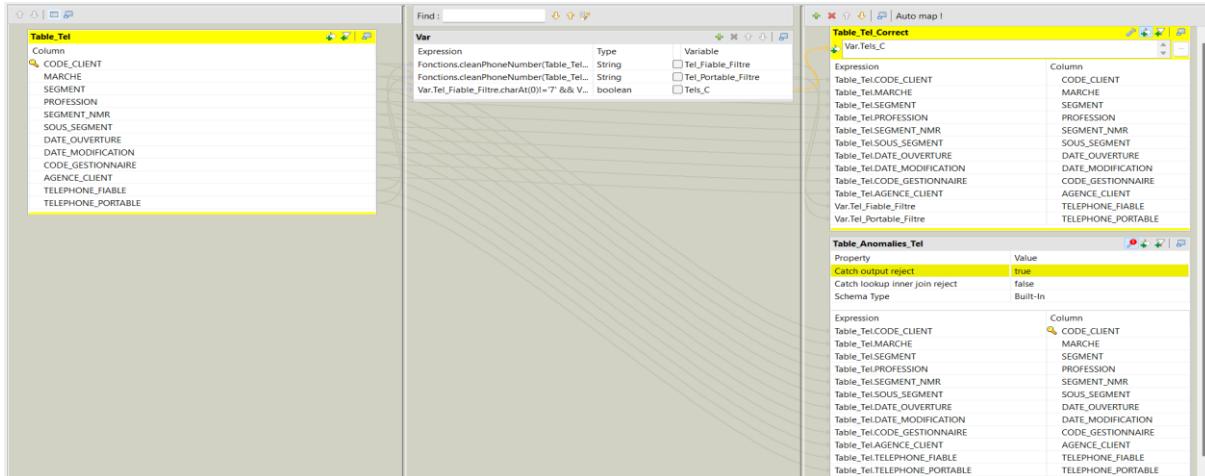


Figure 98 : Composant « tMap_Tel »

La figure 98 présente l'interface de l'éditeur de mapping du composant tMap. À gauche, nous pouvons voir le schéma de la table d'entrée nommées « Tel ». À droite, le schéma des tables

de sortie « Tel_corrects » et « Anomalies_Tel ». Pour extraire les données de cette table d'entrée, il est nécessaire de faire le mapping entre la table d'entrée et les deux tables de sorties. Pour ce faire, il suffit de faire glisser les colonnes de la table « Tel » vers les deux tables de sorties et d'activer l'option ‘catch output reject’ sur ‘True’ pour récupérer les numéros de téléphone qui ne respectent pas les règles de gestion qui sont présent dans les figures de 100 à 102.

Avant d'appliquer les règles de gestion nous allons mettre les numéros de téléphone sous la forme numérique et nous allons vérifier s'ils commencent par un préfixe ou non (par exemple pour les numéros tunisiens le préfixe est « +216 » ou « 216 »). Pour ce faire nous avons créé une fonction en Java dans la routine (figure 99).

```
public static String cleanPhoneNumber(String phoneNumber) {

    // Enlever tous les caractères non numériques
    phoneNumber = phoneNumber.replaceAll("[^0-9]", "");
    if(phoneNumber.length() != 8) {
        String prefix = "216"; // Mettre ici le préfixe à enlever
        if (phoneNumber.startsWith(prefix)) {
            phoneNumber = phoneNumber.substring(prefix.length());
        }
    }
    return phoneNumber;
}
```

Figure 99 : Fonction Java pour standardiser les numéros de téléphone

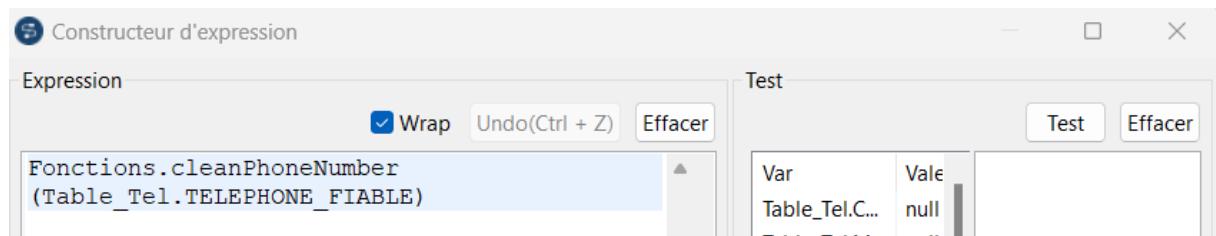


Figure 100 : Application de la fonction « cleanPhoneNumber » pour les numéros de téléphone fiable

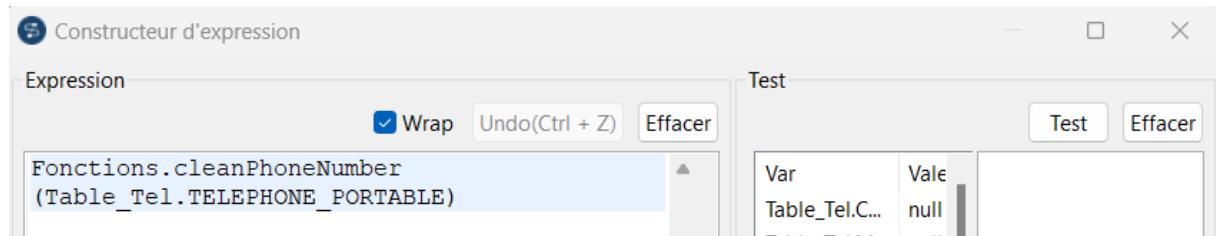


Figure 101 : Application de la fonction « cleanPhoneNumber » pour les numéros de téléphone portable

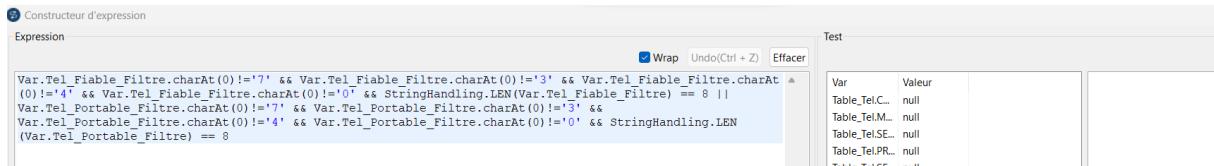


Figure 102 : Expression pour la règle de gestion qui consiste à avoir les numéros de téléphone dont la longueur est exactement 8 chiffres et qui ne commence pas par 0 ou 3ou 4ou bien 7

3. La troisième étape consiste à charger les tables dans la base de données ORACLE. Pour se faire nous avons utilisé le composant **tOracleOutput**. Les figures 103 et 104 montrent la configuration des deux composants **tOracleOutput** pour chaque table.

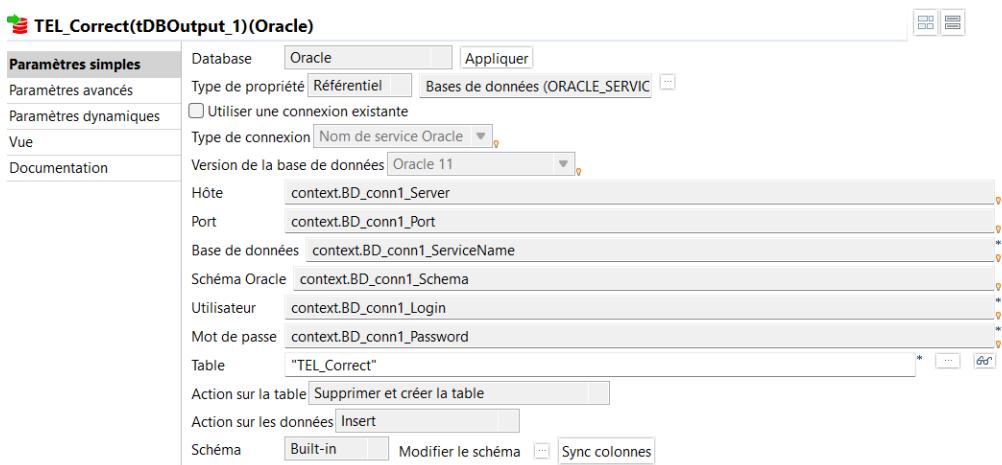


Figure 103 : Composant **tOracleOutput** pour la table « Tel_Corrects »

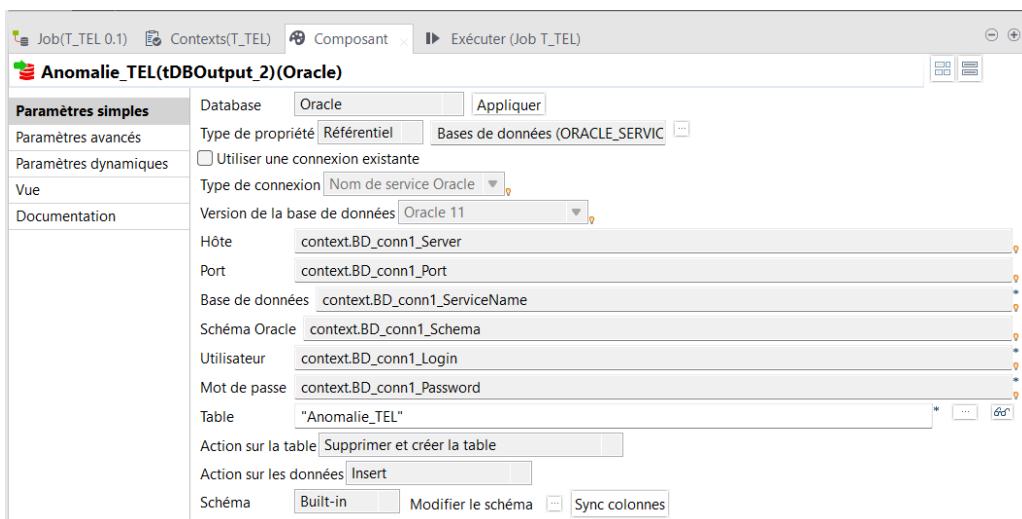


Figure 104 : Composant **tOracleOutput** pour la table « Anomalies_Tel »

Les figures de 105 à 107 montrent l'exécution du Job « T_Tel » (figure 105) ainsi que le chargement des deux tables « Tel_Corrects » et « Anomalies_Tel » dans la base ORACLE (figure 106 et 107).

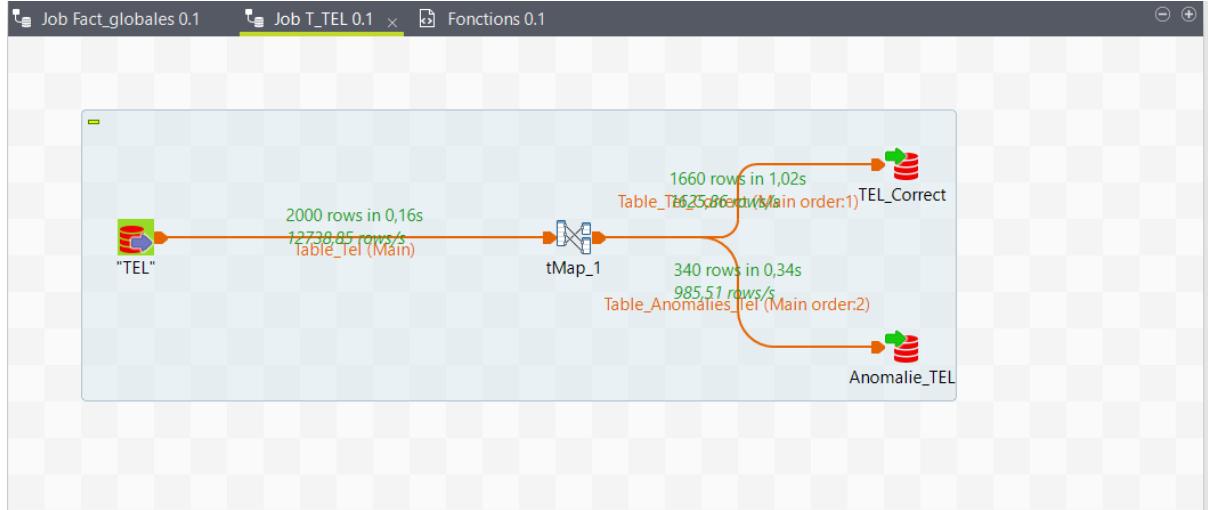


Figure 105 : Exécution du job « T_Tel »

	script_MAJ_Mensuelle.sql	Page de bienvenue	Detect_Anomalies	NOM_PRENOM_Correct
	Colonnes	Données	Modèle	Contraintes
	Tracé	Filtre		
1	ENT [] MARCHE [] PROFESSION [] SPECIFIQUE [] SOUS_SEGMENT [] DATE_OUVERTURE [] DATE_MODIFICATION [] DATE_DERNIERE_MISE_AJOURNEE [] AGENCIE_CLIENT [] NOM [] PRENOM			
2	Enterprise P16 S5 I3	TR GEI 03/02/09 15/03/09 150	219	21435 90 817 8
3	TPE P13 S2 I4	TR GEI 05/03/13 14/04/13 953	27	4 97 5 6006 +21445 8 72351
4	TPE P7 S4 I1	BANQUE PRIVEE 24/07/16 02/05/16 386	9	214350957 238 +21431 8536 11
5	Profession libérale P9 S3 I2	TR GEI 26/05/19 05/07/19 843	104	(null)
6	TPE P12 S5 I4	VIP 20/12/14 29/01/15 944	187	(null) Rahma
7	Entreprise P16 S1 I2	DIGITAL TSE 03/07/17 12/08/17 883	65	(null) Majdi
8	Entreprise P16 S2 I3	BANQUE PRIVEE 08/06/22 18/07/22 483	134	Hajji (null)
9	Profession libérale P10 S1 I5	BANQUE DIGITALE 17/02/10 29/03/10 294	161	(null) Nadhem
10	TPE P7 S5 I1	OFF SHORE 19/07/20 28/08/20 963	232	Othmani (null)
11	Entreprise P16 S2 I5	DIGITAL TSE 04/11/17 14/12/17 724	182	(null) Hanen
12	TPE P14 S5 I3	TR GEI 26/08/17 05/10/18 179	147	Majri Mehdji
13	TPE P15 S2 I4	VIP 06/07/16 15/08/16 403	159	Cherni Abdelhamid
14	TPE P14 S5 I2	VIP 01/05/15 10/06/15 315	94	(null) Maran
15	Profession libérale P9 S4 I5	BANQUE DIGITALE 15/03/18 24/04/18 256	44	(null) Lotfi
16	Profession libérale P11 S4 I4	DIGITAL TSE 01/12/22 10/02/23 454	145	Zribi (null)
17	TPE S6 S5 I1	OFF SHORE 14/07/08 23/08/08 919	5	Bersiche Tora
18	TPE P7 S2 I1	OFF SHORE 20/06/19 07/08/19 514	79	(null) Khalil
19	Profession libérale P11 S4 I1	OFF SHORE 26/03/13 05/05/13 356	235	Toimi Maissem
20	TPE P6 S4 I2	OFF SHORE 21/11/19 31/12/08 981	141	Hammami Kais
21	TPE P14 S5 I2	TR GEI 21/03/10 30/04/10 222	187	(null) Majdi
22	Profession libérale P10 S5 I5	BANQUE PRIVEE 07/09/15 16/09/15 454	12	Mahrouk (null)
23	Profession libérale P10 S3 I3	BANQUE PRIVEE 16/09/15 07/12/15 859	36	Gharbi Mehdji
24	Particuliers P4 S1 I3	BANQUE DIGITALE 05/01/23 14/02/23 500	54	Bouhamed (null)
25	Profession libérale P9 S2 I1	BANQUE PRIVEE 16/07/17 25/08/17 446	61	(null) Ali
26	TPE P6 S3 I4	BANQUE PRIVEE 25/05/08 04/07/08 938	69	Amri Zied
27	Particuliers P3 S2 I4	BANQUE DIGITALE 28/10/15 07/12/15 859	36	Gharbi Mehdji
28	TOP	TO SQL	18/09/99 24/04/23 649	94

Figure 106 : Chargement de la table « Tel_correctes »

	script_MAJ_Mensuelle.sql	Page de bienvenue	Detect_Anomalies	ANOMALIE_TEL
	Colonnes	Données	Modèle	Contraintes
	Tracé	Filtre		
1	ANOMALIE_TEL [] CODE_CLIENT [] MARCHE [] PROFESSION [] SPECIFIQUE [] SOUS_SEGMENT [] DATE_OUVERTURE [] DATE_MODIFICATION [] CODE_GESTIONNAIRE [] AGENCIE_CLIENT [] TELEPHONE_FISIBLE [] TELEPHONE_PORTABLE []			
2	P16 S5 I4 TR GEI 03/02/09 15/03/09 150	10/07/11 26/09/11 953	219	21435 90 817 8 21439 749 856
3	P13 S1 I1 VIP 24/08/08 05/10/08 251	17/09/11 31/10/08 160	27	4 97 5 6006 +21445 8 72351
4	P17 S1 I1 BANQUE DIGITALE 10/03/23 19/03/23 690	222	214350957 238 +21431 8536 11	
5	P14 S1 I2 OFF SHORE 31/01/08 11/03/08 156	12	+21435 5 6 4 59 2 8 21435 5 7 0 0 847	
6	P2 S1 I4 TR GEI 10/02/17 22/03/17 445	223	4 5 815 1 82 +21404 0 858 0 6	
7	P17 S1 I1 OFF SHORE 20/11/10 30/12/10 327	31	+21404 2 15 7 0 3 +21404 2 1973 34 3	
8	P17 S1 I1 BANQUE DIGITALE 24/12/13 04/02/14 574	31	+21404 2 145 69 +2144 2173 34 3	
9	P16 S1 I1 VIP 28/12/10 31/12/10 111	231	+21404 2 145 72 +21404 2 145 75	
10	P2 S4 I4 VIP 13/09/18 22/04/18 837	167	+21433 0 6 57 7 5 0 4 3 2 433 3 +21433 0 6 57 7 5 0 4 3 2 433 3	
11	P7 S3 I3 DIGITAL TSE 23/05/18 05/06/18 804	85	+21404 2 145 72 +21404 2 145 75 +21433 0 6 57 7 5 0 4 3 2 433 3 +21433 0 6 57 7 5 0 4 3 2 433 3	
12	P3 S2 I3 TR GEI 01/07/18 10/08/18 554	149	+21400 02/07/82 21604487 098	
13	P6 S5 I5 BANQUE DIGITALE 16/06/09 24/07/09 171	234	+214452 5 4 5 3 9 2160 1 2527 80	
14	P4 S3 I1 BANQUE DIGITALE 27/04/17 06/06/17 135	66	+214390 92 162 +21400 6 3 0 0 11	
15	P12 S5 I5 BANQUE DIGITALE 02/09/10 12/10/10 845	113	0 28163 52 +2160 4 3 1 2739	
16	P14 S5 I5 BANQUE PRIVEE 10/01/11 19/02/11 180	154	41 8 0 91 8 4 +21471 42355 5	
17	P12 S1 I3 BANQUE DIGITALE 11/04/19 21/07/19 453	22	+214390 556 3 4 1075 23 8	
18	P16 S4 I4 BANQUE PRIVEE 09/11/08 09/11/08 540	68	3 1 7269 0 +2140 6 33 2097	
19	P12 S4 I5 VIP 05/02/11 17/03/11 880	77	+2160 2 9 4 6 2 25 2160 4 351 25 2	
20	P17 S4 I3 BANQUE PRIVEE 03/09/18 12/09/18 529	31	+21404072 4 2 7 2164 9 7 7236	
21	P16 S2 I4 VIP 13/10/14 22/11/14 887	13	2160 2 7420 2 8 2167 2 60 323 4	
22	P4 S3 I5 VIP 17/11/11 27/12/12 110	173	+2140051 8 5 7 3 9 8861 0	
23	P3 S5 I3 TR GEI 25/12/11 03/02/13 636	44	+2144 9 5 38 23 8 2163 7 3759 57	
24	P2 S1 I3 BANQUE DIGITALE 02/09/18 12/10/18 906	218	4 2 51 347 +21432 4333 8 1	
25	P9 S4 I3 DIGITAL TSE 05/06/18 18/07/18 346	59	+2144 6733279 7 13001 96	
26	P12 S3 I4 BANQUE DIGITALE 16/06/19 02/07/19 982	76	+214390 52256 9 2140 6 33 2097 8 4	
27	P16 S5 I3 BANQUE PRIVEE 20/08/18 07/07/18 443	133	0 2816 4 6 2 25 2160 4 3 1 2739 0 5 55 14 8 6	
28	TOP	TOP	21/09/99 24/04/23 649	94

Figure 107 : Chargement de la table « Anomalies_Tel »

2.3.6 Application des règles de gestion pour transformer les données liées aux adresses mails des clients

La figure 108 montre l'environnement du job concernant les transformations effectuées sur les données importées de la table « EMAIL » qui a été chargée dans l'étape précédente.

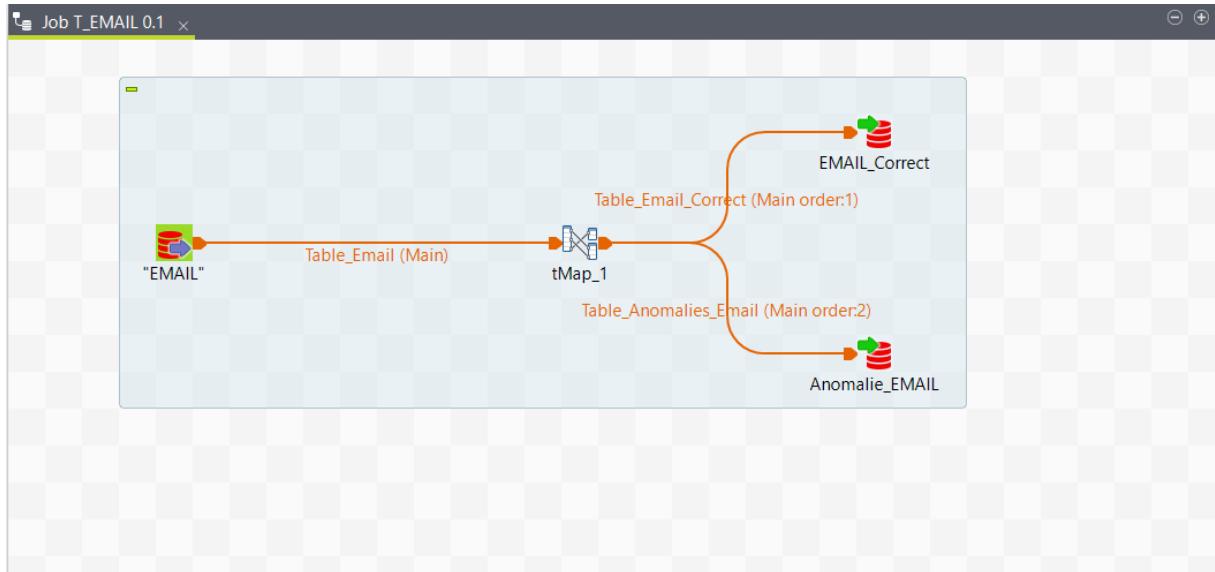


Figure 108 : Environnement de travail du job « T_Email »

Ce job permet d'appliquer les règles de gestion pour les adresses mail des clients. Une adresse mail est correcte si :

- Elle ne contient pas des symbole (exemple : *, # ; / ...)
- Elle se termine par un service mail (exemple : @gmail.com, @yahoo.fr ...)
- Elle commence par un caractère alphabétique

1. La première étape pour la réalisation de ce job consiste à importer la table Tel en utilisant le composant d'entrée **tOracleInput**. Les figures 109 et 110 montrent ce composant ainsi que son schéma.



Figure 109 : Table « Email »

Schéma de "EMAIL"

"EMAIL"

Colonne utilisée	Colonne	Db Column	Clé	Type	Type ...	N.	Modèle de ...	Long...	Précis...	Par ...	Com...
✓	CODE_CLIENT	CODE_CLIENT	✓	BigD...	NUM...			38	0		
✓	MARCHE	MARCHE		String	VARC...	✓		20	0		
✓	SEGMENT	SEGMENT		String	VARC...	✓		20	0		
✓	PROFESSION	PROFESSION		String	VARC...	✓		20	0		
✓	SEGMENT_NMR	SEGMENT_NMR		String	VARC...	✓		20	0		
✓	SOUS_SEGMENT	SOUS_SEGMENT		String	VARC...	✓		20	0		
✓	DATE_OUVERTURE	DATE_OUVERTURE		Date	DATE	✓	"dd/MM/y...	20	0		
✓	DATE_MODIFICATION	DATE_MODIFICATION		Date	DATE	✓	"dd/MM/y...	20	0		
✓	CODE_GESTIONNAIRE	CODE_GESTIONNAIRE		BigD...	NUM...	✓		38	0		
✓	AGENCE_CLIENT	AGENCE_CLIENT		BigD...	NUM...	✓		38	0		
✓	EMAIL	EMAIL		String	VARC...	✓		50	0		

OK Cancel

Figure 110 : Structure de la table « Email »

2. La deuxième étape consiste à l'utilisation du composant **tMap** pour récupérer les numéros de téléphones corrects ainsi que ceux présentant des anomalies. La figure 111 montre ce composant.

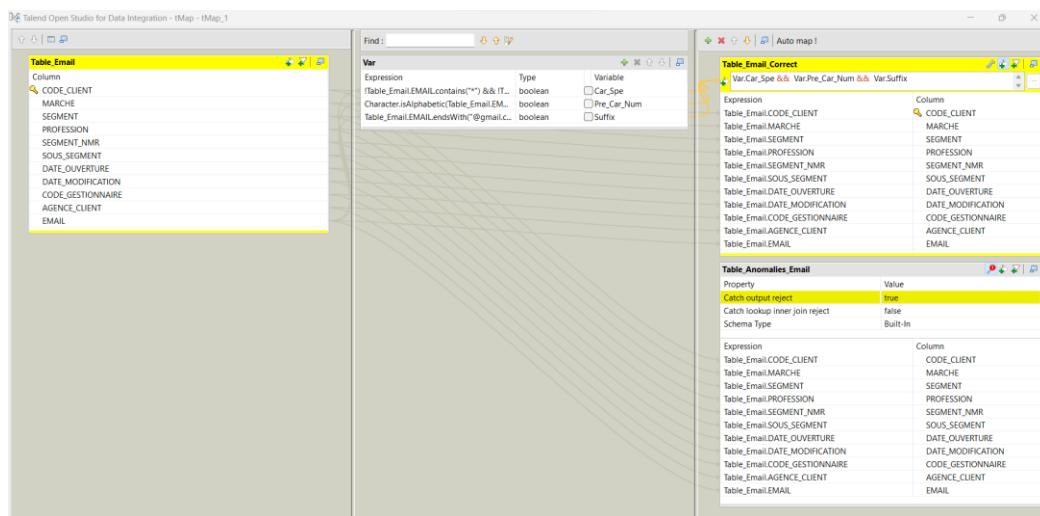


Figure 111 : Composant « tMap_Mail »

La figure 111 présente l'interface de l'éditeur de mapping du composant **tMap**. À gauche, nous pouvons voir le schéma de la table d'entrée nommées « Tel ». À droite, le schéma des tables de sortie « Email_corrects » et « Anomalies_Email ». Pour extraire les données de cette table d'entrée, il est nécessaire de faire le mapping entre la table d'entrée et les deux tables de sorties. Pour ce faire, il suffit de faire glisser les colonnes de la table « Email » vers les deux tables de sorties et d'activer l'option ‘catch output reject’ sur ‘True’ pour récupérer les adresses mail qui ne respectent pas les règles de gestion qui sont présent dans les figures de 112 à 114.

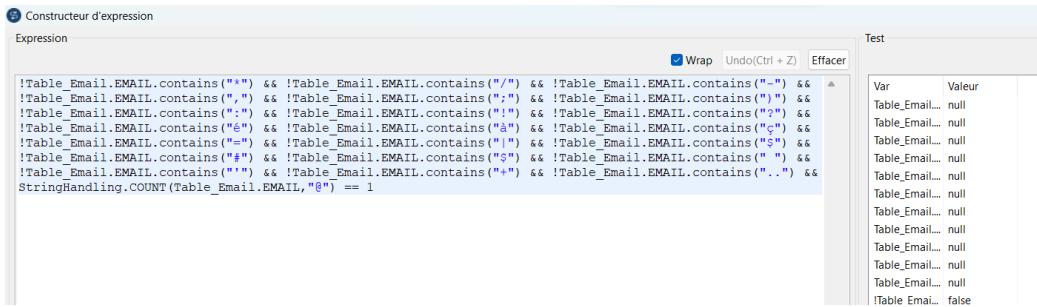


Figure 112 : Expression pour la règle de gestion qui consiste à avoir les adresses mail qui ne contiennent pas des caractères spéciaux



Figure 113 : Expression pour la règle de gestion qui consiste à avoir les adresses mail qui contiennent un nom de service mail



Figure 114 : Expression pour la règle de gestion qui consiste à avoir les adresses mail qui commence avec un caractère alphabétique

3. La troisième étape consiste à charger les tables dans la base de données ORACLE. Pour se faire nous avons utilisé le composant **tOracleOutput**. Les figures 115 et 116 montrent la configuration des deux composants **tOracleOutput** pour chaque table.

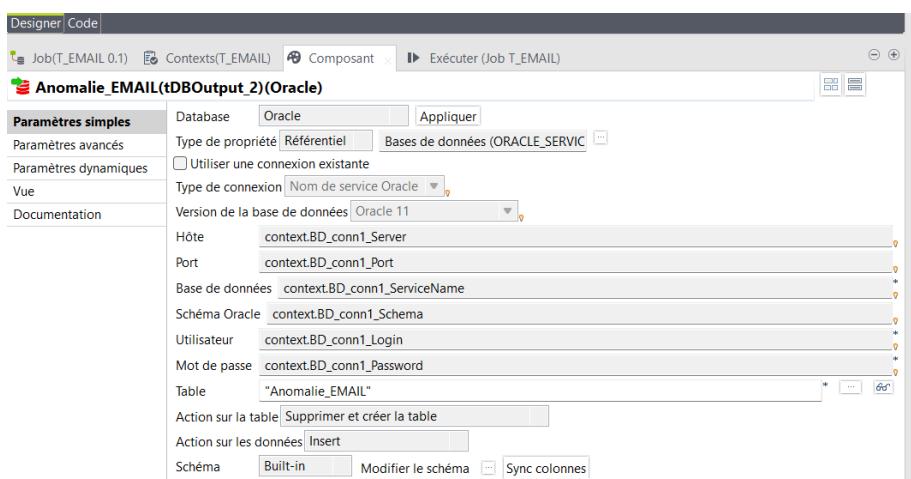


Figure 115 : Composant **tOracleOutput** pour la table « Anomalies_Email »

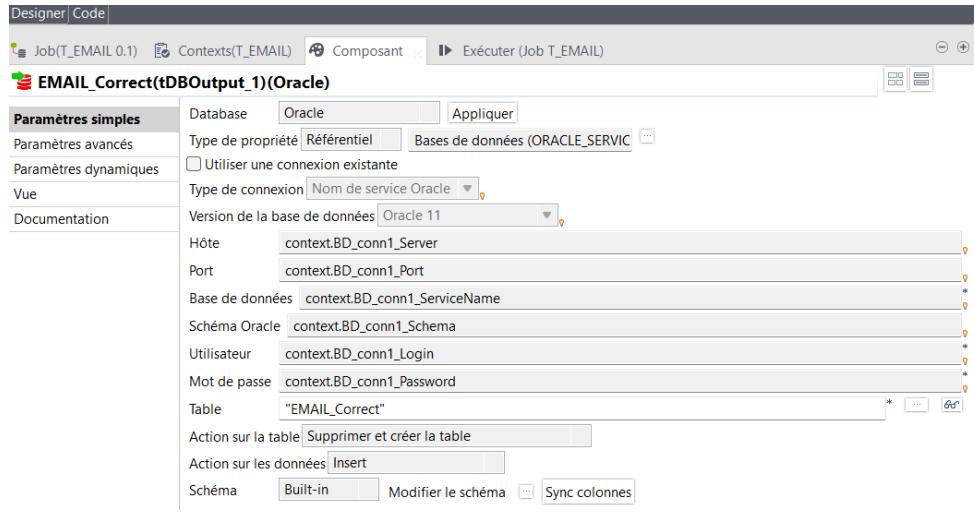


Figure 116 : Composant tOracleOutput pour la table « Email_Corrects »

Les figures de 117 à 119 montrent l'exécution du Job « T_Email » (figure 117) ainsi que le chargement des deux tables « Email_Corrects » et « Anomalies_Email » dans la base ORACLE (figure 118 et 119) :

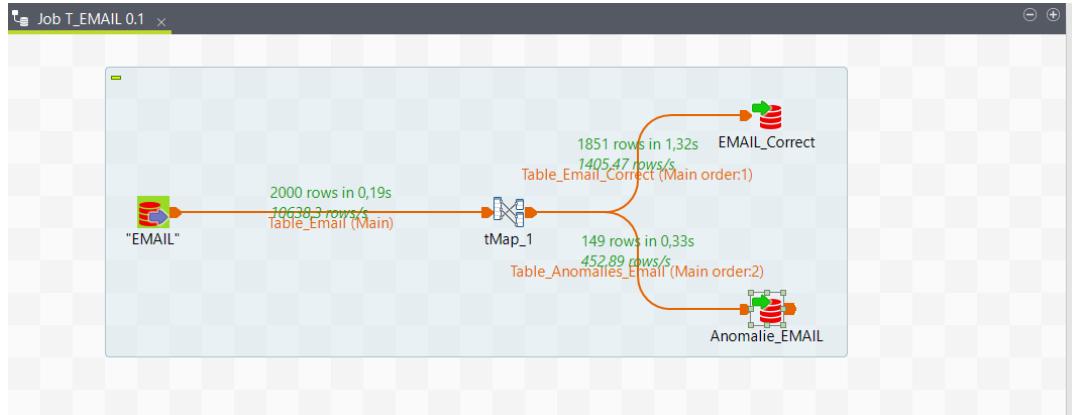


Figure 117 : Exécution du job « T_Email »

The screenshot shows the Oracle Database SQL Developer interface with the following details:

- Connections:** script_MAJ_Historique
- Table:** EMAIL_CORRECT
- Columns:** PROFESSION, SEGMENT_NMR, SOUS_SEGMENT, SPECIFIQUE, DATE_OUVERTURE, DATE_MODIFICATION, CODE_GESTIONNAIRE, AGENCE_CLIENT, EMAIL
- Data:** The table contains approximately 1851 rows of data, with columns showing various email metadata and recipient information.

Figure 118 : Chargement de la table « Email_corrects »

	PROFESSION	SEGMENT_NMR	SOUS_SEGMENT	DATE_OUVERTURE	SPECIFICITE	DATE_MODIFICATION	CODE_GESTIONNAIRE	AGENCE_CLIENT	EMAIL
1	P10	S1	14	10/04/17	TR GEI	2/07/17	612	16	Ben.Ali.Mehdi123@hotmail.fr
2	P12	S1	24	02/10/17	VIP	1/1/21/17	611	30	Ben.Ammar.Nabil123@yahoo.com
3	P17	S5	23	08/05/12	TR GEI	1/0/6/12	977	66	Mari.Abdelhamid030310@yahoo.fr
4	P17	S5	21	18/04/11	VIP	0/0/4/23	847	130	Abdeljellil.Omariyousfi12@gmail.com
5	P17	S5	14	04/04/17	VIP	0/0/4/23	478	94	Ammar.Mohktar177p@outlook.com
6	P17	S4	24	29/03/14	OFF SHORE	0/0/5/14	722	136	Ben.Ali.Jabdehainidogou2@hotmail.fr
7	P15	S2	15	19/03/20	TR GEI	2/0/4/20	28	119	Ben.Ali.Baniasrc20@gmail.com
8	P13	S2	15	27/04/12	VIP	0/0/6/12	853	96	Reiji.Omariyousfi178@hotmail.com
9	P12	S1	14	26/05/11	DIGITAL TSF	0/0/7/11	875	97	Ben.Ali.Jabdehainidogou2@hotmail.fr
10	P15	S3	14	11/06/16	DIGITAL TSF	2/1/0/16	617	126	Ben.Ali.Mehdi123@yahoo.fr
11	P4	S3	15	28/09/09	OFF SHORE	0/1/1/09	476	23	Ben.Ammar.Yosra11inf@hotmaill.com
12	P4	S4	14	13/09/09	OFF SHORE	2/3/1/09	280	1	Ben.Ammar.Amenirrgew@outlook.com
13	liberale_P10	S1	14	03/09/19	BANQUE DIGITALE 1/3/0/13	231	48	Ben.Ali.Salahedn49@hotmail.fr	
14	P2	S5	14	14/09/17	DIGITAL TSF	2/1/0/13	645	73	Ben.Ali.Mehdi123@yahoo.fr
15	P3	S5	22	27/03/22	VIP	0/0/5/22	978	125	Ben.Ammar.Mehdi123@yahoo.fr
16	P7	S2	14	08/02/09	BANQUE DIGITALE 2/0/0/19	828	12	Ben.Ali.Mehdi123@yahoo.fr	
17	P12	S1	13	09/11/15	DIGITAL TSF	1/9/12/15	645	5	Ben.Ali.Mehdi123@yahoo.fr
18	P8	S5	15	11/03/18	BANQUE DIGITALE 2/0/0/18	248	83	Ben.Ammar.Neihilb3nor@hotmail.com	
19	P17	S4	11	02/04/13	TR GEI	1/2/0/13	761	79	Ben.Ammar.Kaislahni12@hotmail.fr
20	P8	S3	12	29/06/12	DIGITAL TSF	0/0/8/12	633	138	Ben.Ali.Samirhewi@hotmail.fr
21	P16	S1	11	14/04/10	BANQUE DIGITALE 2/4/0/10	817	167	Ben.Ammar.Ahmeden77m@outlook.com	
22	P4	S3	12	27/04/13	VIP	0/0/6/13	794	185	Ben.Ammar.Amenirrgew@outlook.com
23	P4	S4	14	31/05/11	BANQUE PRIVEE	1/0/7/11	847	82	Ben.Ammar.Samirg11f@outlook.com
24	P3	S1	13	26/04/08	BANQUE PRIVEE	0/0/4/08	716	228	Ben.Ali.Mehdi123@yahoo.fr
25	P1	S5	15	07/11/18	TR GEI	1/7/12/18	817	136	Ben.Ali.Mokhtarjr123@hotmail.fr
26	liberale_P11	S1	14	28/04/15	VIP	0/0/4/15	546	39	Ben.Ammar.Yasminene1end@gmail.com
27	liberale_P10	S4	22	13/08/10	BANQUE DIGITALE 2/2/0/10	382	37	Ben.Ali.Aliledel124@hotmail.com	
28	liberale_P11	S1	49	24/07/18	RESERVE OFFICE	1/1/1/18	441	144	Ben.Ammar.Yasminene1end@gmail.com

Figure 119 : Chargement de la table « Anomalies_Email »

2.4. Charger les dimensions et les tables de faits

Après avoir chargé les données des fichiers sources et appliquer les règles de gestions pour détecter les anomalies, nous pouvons charger nos dimensions et nos tables de faits.

1-Chargement de la dimension Gestionnaire qui contient les informations liées aux gestionnaires de la banque (figure 120) :

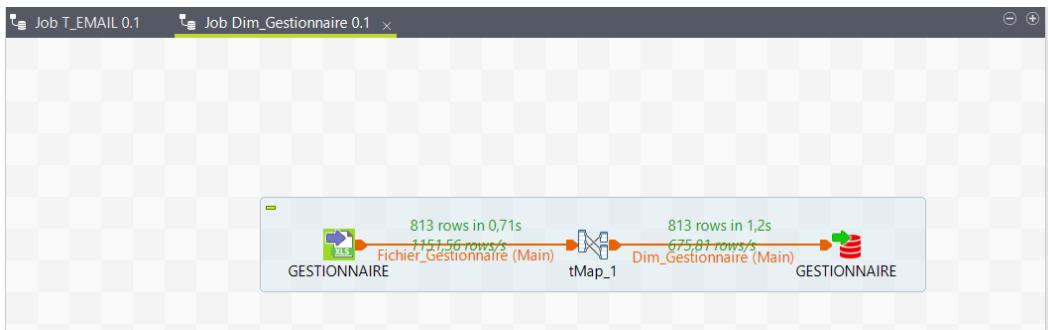


Figure 120 : Chargement de la dimension « Dim_Gestionnaire » à partir du fichier source

2-Chargement de la dimension « Dim_Agence » qui contient les informations des agences de la banque (figure 121) :

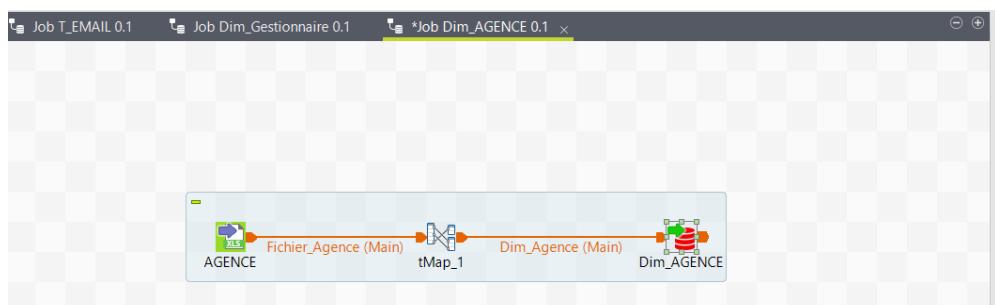


Figure 121 : Chargement de la dimension « Dim_Agence » à partir du fichier source

3-Chargement de la dimension « Dim_Segmentation_clients » qui contient les informations des marchés, profession et segment de la banque (figure 122) :

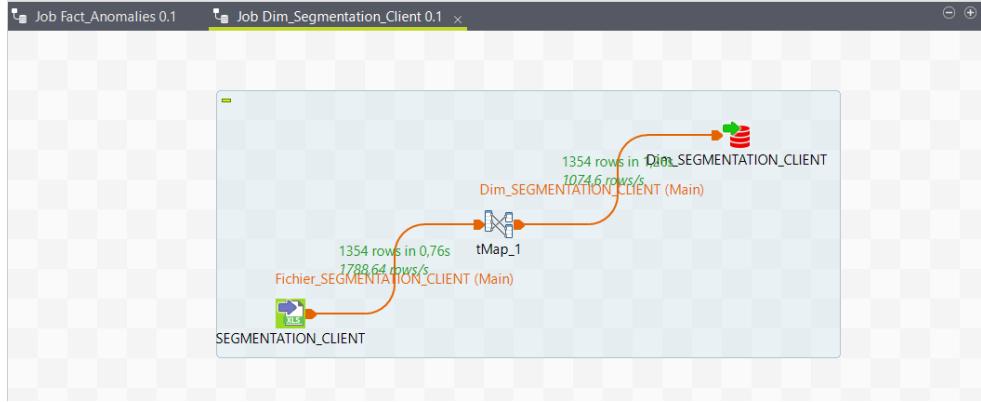


Figure 122 : Chargement de la dimension « Dim_Segmentation_Client » à partir du fichier source

4-Chargement de la dimension « Dim_Client_Anomalies » qui contient les informations des clients de la banque (figure 123) :

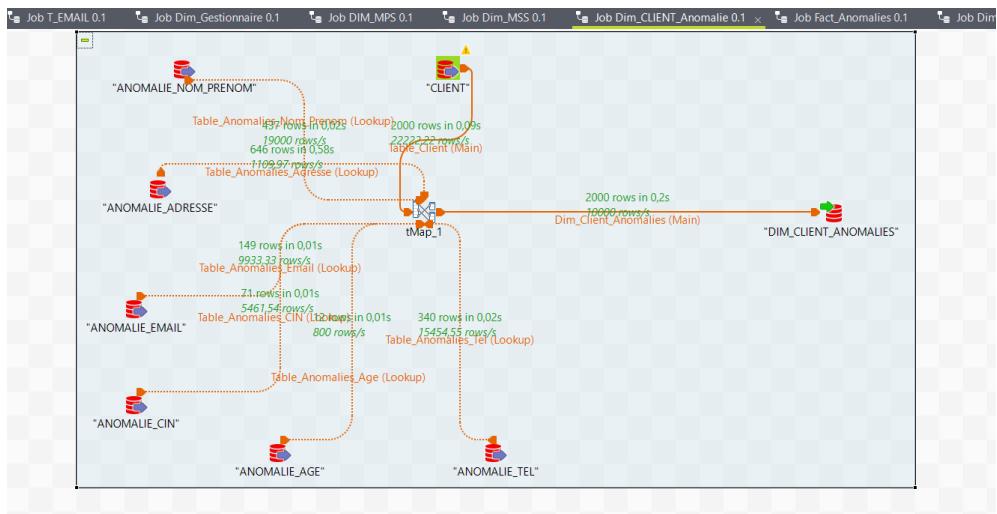


Figure 123 : Chargement de la dimension « Dim_Client_Anomalies »

Ces données seront traitées dans le composant **tMap** et transformées. Nous appliquons la jointure entre la table Client, qui contient toutes les informations des clients correctes et incorrectes, et chaque table anomalie avec le code client qui est commun entre eux.

Le traitement appliqué sur les données consiste à prendre la valeur de chaque use case si la jointure pointe sur une ligne non null dans la table d'anomalie de cette use case (c'est-à-dire elle est incorrecte) sinon, si nous avons une valeur null, nous affectons à la colonne correspondante dans la dimension « dim_client_anomalie » une chaîne « null » pour pouvoir filtrer les lignes présentant des anomalies à ceux qui sont correctes tout en gardant le nombre total des fiches clients.(figure 124)

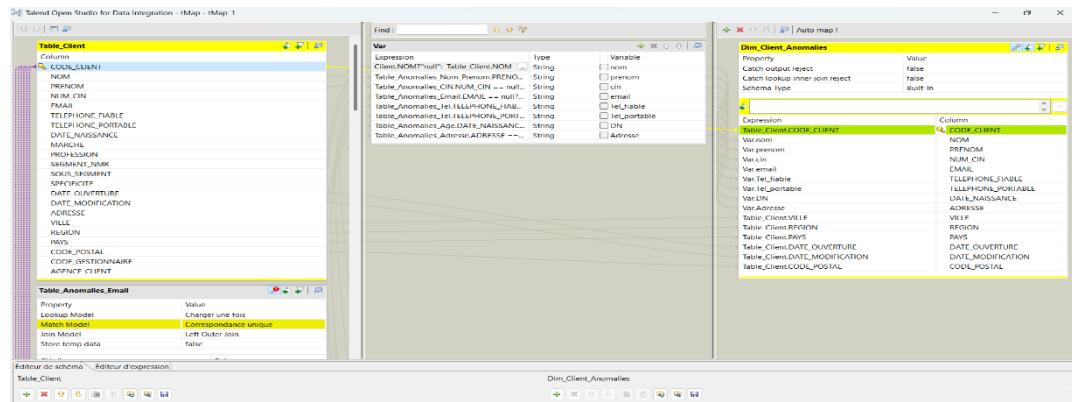


Figure 124 : Composant « tMap_Dim_Client_anomalies »

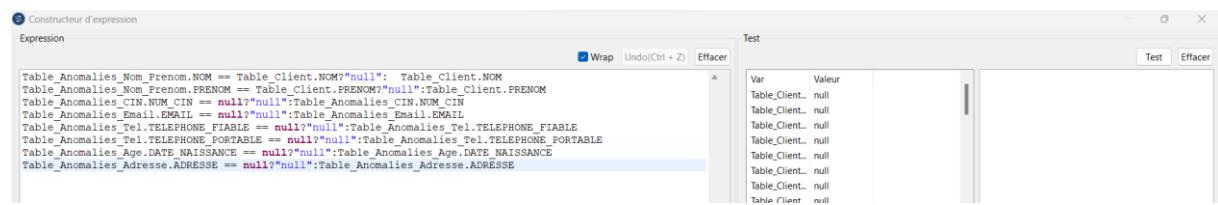


Figure 125 : Expression pour récupérer les données incorrectes et transformer les données correctes en « null »

Chaque ligne de la figure 125 est affectée à une variable de type chaîne. Ces variables vont être affectées aux colonnes correspondantes au niveau de la dimension « Dim_Client_anomalies. »

5-Chargement de la table de fait « Fact_Anomalies » qui va contenir les clés étrangères des dimensions ainsi que les mesures que nous allons créer dans Power BI(figure 126) :

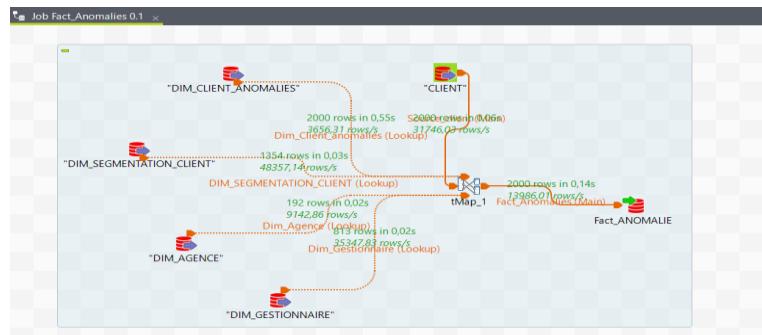


Figure 126 : Chargement de la table des faits « Fact_Anomalies »

Finalement et après avoir créé toutes les tables, les dimensions et la table de fait, nous avons opté à appliquer la notion de « job master » pour regrouper tous les jobs de chaque dossier en un seul.(figure 127, 128 et 129)

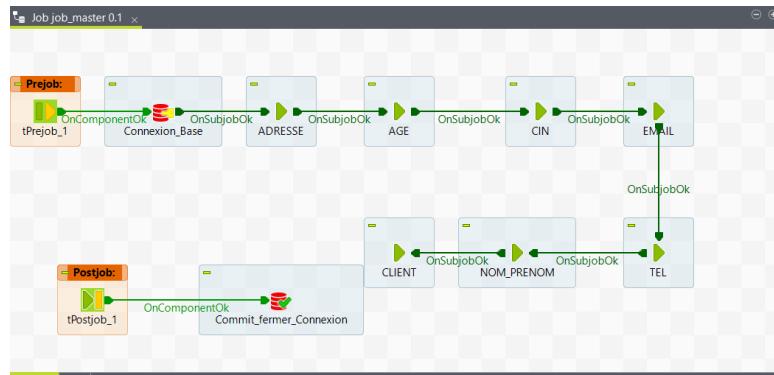


Figure 127 : Job master pour l'ODS

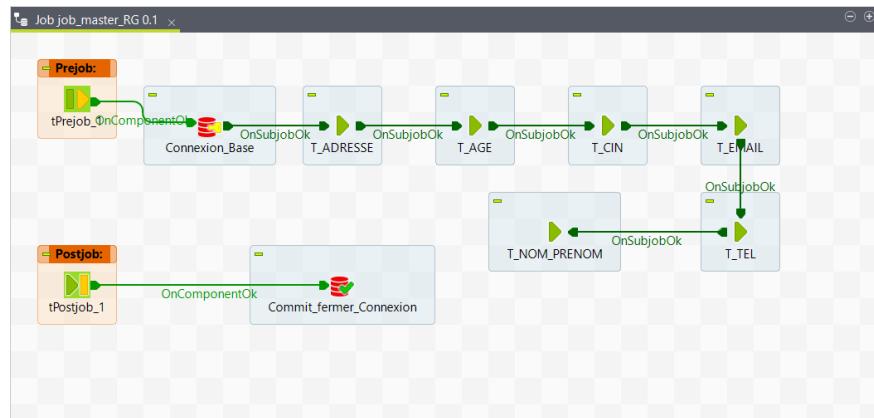


Figure 128 : Job master pour les règles de gestion

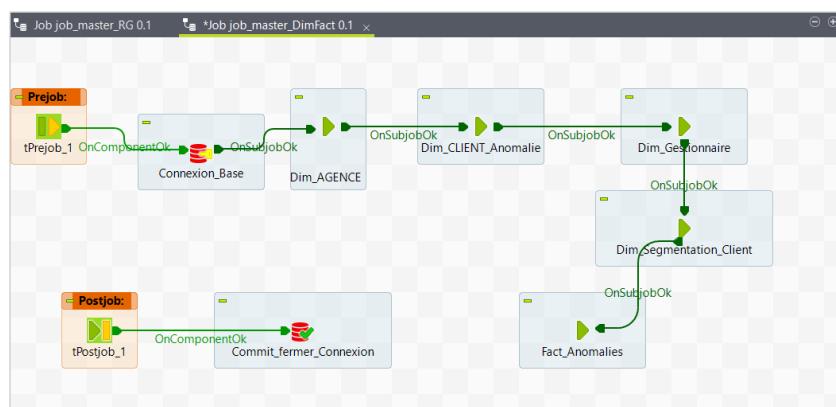


Figure 129 : Job master pour les dimensions et de la table de fait

La dernière étape consiste à définir les clés étrangères de la table des faits à l'aide de Oracle Sql Developer (figure 130) :

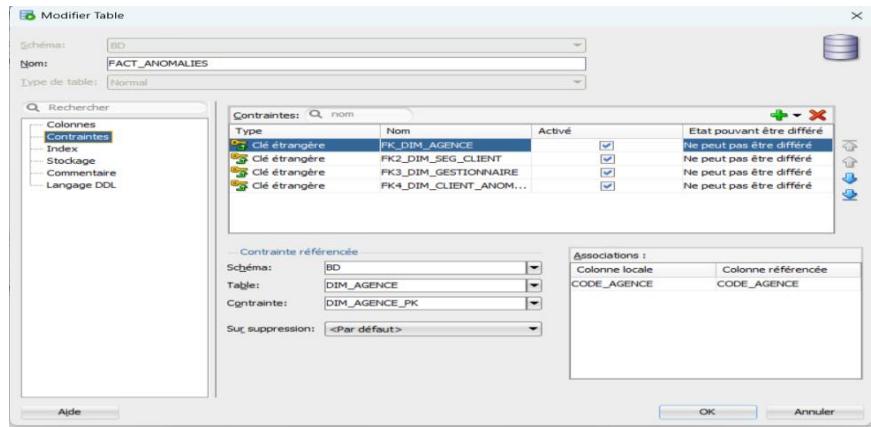


Figure 130 : Clés étrangères de la table des faits

3.Conclusion

Ce troisième chapitre décrit les différentes étapes d'un processus ETL, en commençant par l'extraction des données des fichiers sources et leur chargement dans l'ODS. Ensuite, l'application des règles de gestion pour détecter les anomalies dans les données clientèles. Enfin, le chargement de ces données dans les dimensions en utilisant les tables créées lors de l'application des règles de gestion. Tout cela a été précédé par la simulation de données fictives à l'aide du langage Python.

Chapitre IV. Restitution des données

Introduction

Dans ce quatrième chapitre, nous allons entamer la deuxième partie du projet décisionnel, à savoir la récupération des données stockées dans l'entrepôt de données. Notre mission dans le cadre de ce projet de recherche était de collecter un maximum de données clients afin d'identifier les anomalies de ces données et leur diffusion. Pour cela, nous avons réalisé un rapport analytique sous forme de Storytelling avec l'outil de visualisation de données Power BI Desktop. Ce rapport comprend une page d'accueil ainsi que des pages correspondantes montrant la répartition des données clients, les anomalies ainsi que l'évolution des anomalies dans le temps.

1. Connexion à la base de données, calcul des mesures de la table de faits et vérification de la structure du modèle de données

La première étape à réaliser est l'obtention des données à manipuler. Pour cela, il suffit d'ouvrir l'onglet « Obtenir les données » et de sélectionner la source de notre entrepôt de données. Dans notre cas, nous devons choisir l'option « Base de données Oracle » comme nous pouvons le voir dans la figure 131.

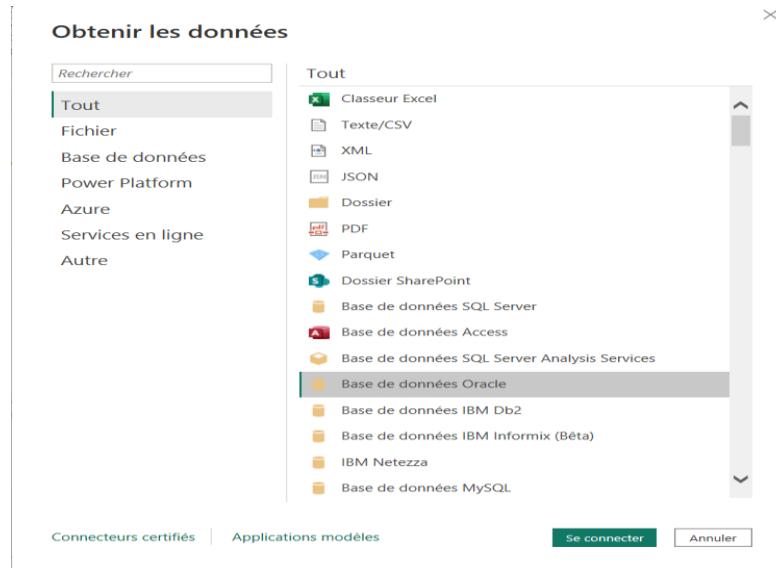


Figure 131 : Importation des données du Data Warehouse

Par la suite, il est nécessaire de remplir les informations de notre base de données afin de configurer la connexion à celle-ci, comme illustré dans la figure 132.

CODE_MPS	CODE_MSS	CODE_AGENCIE	CODE_GESTIONNAIRE	CODE_CLIENT	BD
9	6	124	331	778	
8	5	220	709	779	
12	5	160	960	780	
23	16	231	410	781	
21	13	170	798	782	
11	7	102	261	783	
4	3	171	459	784	
24	14	115	487	785	
16	11	198	528	786	
1	1	191	534	787	
20	16	96	196	788	
7	8	221	105	789	
17	10	104	481	790	
4	2	197	822	791	
2	1	184	301	792	
11	6	108	396	793	
15	10	145	843	794	
16	12	42	102	795	
22	14	1	732	796	
15	12	113	124	797	
8	7	106	648	798	
22	13	175	257	799	

Figure 132: Configuration de la connexion à la base de données Oracle

Une fois les données importées de notre base de données, il est nécessaire de créer les mesures de notre table de faits, « Fact_Anomalies », conformément à la conception de notre Data Warehouse. Le tableau 4 présentent les formules DAX utilisées pour la création de ces mesures.

Tableau 4 : Formules DAX des mesures dans Power BI

Mesures	Descriptions	Formules DAX
-nbr_fiche_client_correctes	-Cette mesure permet de calculer le nombre de fiches client correctes.	<pre> nbr_fiche_client_correctes = COUNTRROWS(FILTER(DIM_CLIENT_ANOMALIES, DIM_CLIENT_ANOMALIES[DATE_NAISSANCE] == "null" && DIM_CLIENT_ANOMALIES[ADRESSE] == "null" && DIM_CLIENT_ANOMALIES[EMAIL] == "null" && DIM_CLIENT_ANOMALIES[NUM_CIN] == "null" && DIM_CLIENT_ANOMALIES[Nom et prenom] <> "null" && DIM_CLIENT_ANOMALIES[Tel] == "1")) </pre>
- nbr_fiche_anomalies	-Cette mesure permet de calculer le nombre de fiches ayant des anomalies.	<pre> nbr_fiche_anomalies = COUNTRROWS(FACT_ANOMALIES) - [nbr_fiche_client_correctes] </pre>
-Profession_Plus_Anomalies_AD	<p>-Cette mesure renvoie la profession ayant le plus d'anomalies pour les adresses.</p> <p>Remarque :</p> <p>A chaque fois en change le type d'anomalie (entre email, nom et prénom, téléphone, numéros de CIN et date de naissance) et nous créons 6 mesures pour chacun.</p>	<pre> Profession_Plus_Anomalies_AD = VAR MaxValue = MAX(T_nbr_adresse_profession[Valeur]) RETURN CALCULATE(FIRSTNONBLANK(T_nbr_adresse_profession[PROFESSION], 1), T_nbr_adresse_profession[Valeur] = MaxValue) </pre>
- Segment_Plus_Anomalies_AD	<p>-Cette mesure renvoie le segment NMR ayant le plus d'anomalies pour les adresses. A chaque fois en change le type d'anomalie (entre email, nom et prénom, téléphone, numéros de CIN et date de naissance) et nous</p>	<pre> Segment_Plus_Anomalies_AD = VAR MaxValue = MAX(T_nbr_adresse_segment[Valeur]) RETURN CALCULATE(FIRSTNONBLANK(T_nbr_adresse_segment[SEGMENT_NMR], 1), T_nbr_adresse_segment[Valeur] = MaxValue) </pre>

	créons 6 mesures pour chacun.	
- Marche_Plus_Anomalies_AD	-Cette mesure renvoie le marché ayant le plus d'anomalies pour les adresses. A chaque fois en change le type d'anomalie (entre email, nom et prénom, téléphone, numéros de CIN et date de naissance) et nous créons 6 mesures pour chacun.	Marche_Plus_Anomalies_AD = VAR MaxValue = MAX(T_nbr_adresse_marche[Valeur]) RETURN CALCULATE(FIRSTNONBLANK(T_nbr_adresse_marche[MARCHE], 1), T_nbr_adresse_marche[Valeur] = MaxValue)
- Taux_anomalies_Adresse	-Cette mesure permet de calculer le taux des anomalies adresse par rapport au nombre total des anomalies. A chaque fois en change le type d'anomalie (entre email, nom et prénom, téléphone, numéros de CIN et date de naissance) et nous créons 6 mesures pour chacun.	Taux_anomalies_Adresse = COUNTRROWS(FILTER(DIM_CLIENT_ANOMALIES, DIM_CLIENT_ANOMALIES[ADRESSE] <> "null"))/[nbr_fiche_anomalies]
Taux_anomalies_Adresse_total	-Cette mesure permet de calculer le taux des anomalies adresse par rapport au nombre total des fiches clients. A chaque fois en change le type d'anomalie (entre email, nom et prénom, téléphone, numéros de CIN et date de naissance) et nous	Taux_anomalies_Adresse_total = COUNTRROWS(FILTER(DIM_CLIENT_ANOMALIES, DIM_CLIENT_ANOMALIES[ADRESSE] <> "null"))/COUNTRROWS(FACT_ANOMALIES)

	créons 6 mesures pour chacun.	
-Taux_anomalies	-Cette mesure permet de calculer le taux des anomalies par rapport au nombre de fiches clientèle total.	Taux_anomalies = [nbr_fiche_anomalies]/COUNTROWS(FACT_ANOMALIES)

La figure 133 montre le modèle en étoile de notre Data Warehouse dans l'outil Power BI.

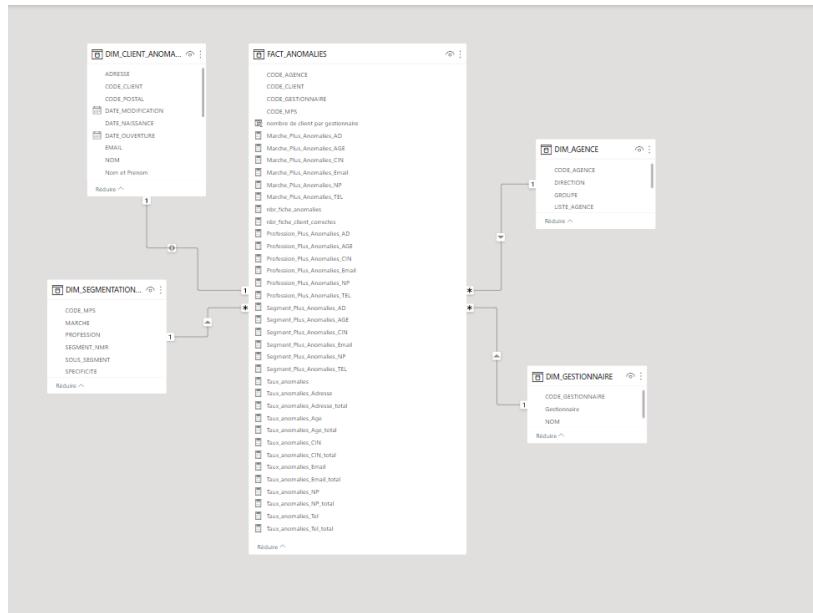


Figure 133 : Modélisation en étoile du Data Warehouse dans Power BI

2. Page d'accueil du rapport

Rapport sur les anomalies des données clientèles de la banque

Objectif et structure du rapport:

Ce Dashboard aide les décideurs à avoir une vue globale sur les anomalies des données clientèle au sein de la banque. Le rapport s'organise comme suit:

- 1)Page d'accueil:** Fournit un aperçu des données clientèles, telles que le nombre total de fiches clients, le nombre de fiches clients correctes, la répartition des fiches clients selon les agences, les marchés et les segments.
- 2)Informations générales:** Fournit un aperçu des données clientèles, telles que le nombre total de fiches clients, le nombre de fiches clients correctes, la répartition des fiches clients selon les agences, les marchés et les segments.
- 3)Anomalies:** Présente une vue globale sur les anomalies des données clientèles, notamment le nombre de fiches clients ayant des anomalies, le taux des fiches clients ayant des anomalies et la répartition des fiches clients ayant des anomalies par gestionnaires, agence. En plus elle permet de voir les anomalies spécifique pour chaque use case, notamment pour les anomalies liées aux adresses, numéros de téléphone, numéros de CIN, date de naissance et noms et prénoms. Elle présente le nombre de fiches clients ayant des anomalies, le taux des fiches clients ayant des anomalies ,le marché, le segment et la profession ayant le plus de fiche client avec des anomalies pour chaque use case.
- 4)Evolution des anomalies au cours du temps:** Donne un aperçu des anomalies des données clientèles au fil du temps pour chaque use case, notamment pour les anomalies liées aux adresses, numéros de téléphone, numéros de CIN, date de naissance et noms et prénoms.

Figure 134 : Page d'accueil du rapport

La figure 134 illustre la page d'accueil de notre rapport analytique. Cette page vise à introduire le sujet d'analyse étudié et permet à l'utilisateur de comprendre l'objectif et la structure du rapport grâce à un texte explicatif présentant les différents thèmes abordés au cours de cette phase de reporting. Nous avons pris soin de respecter la charte graphique d'Attijari Bank en utilisant ses couleurs et en incluant son logo sur toutes les pages de notre rapport. De plus, nous avons intégré un menu de navigation afin de faciliter l'expérience utilisateur en permettant une transition fluide d'une page du rapport à une autre.

3. Analyse et interprétation des graphiques de la détection des anomalies des données clientèles de la banque

Continuons maintenant en nous focalisant sur la première page de notre rapport, qui donne une vision globale sur la répartition des données clientèle de la banque.

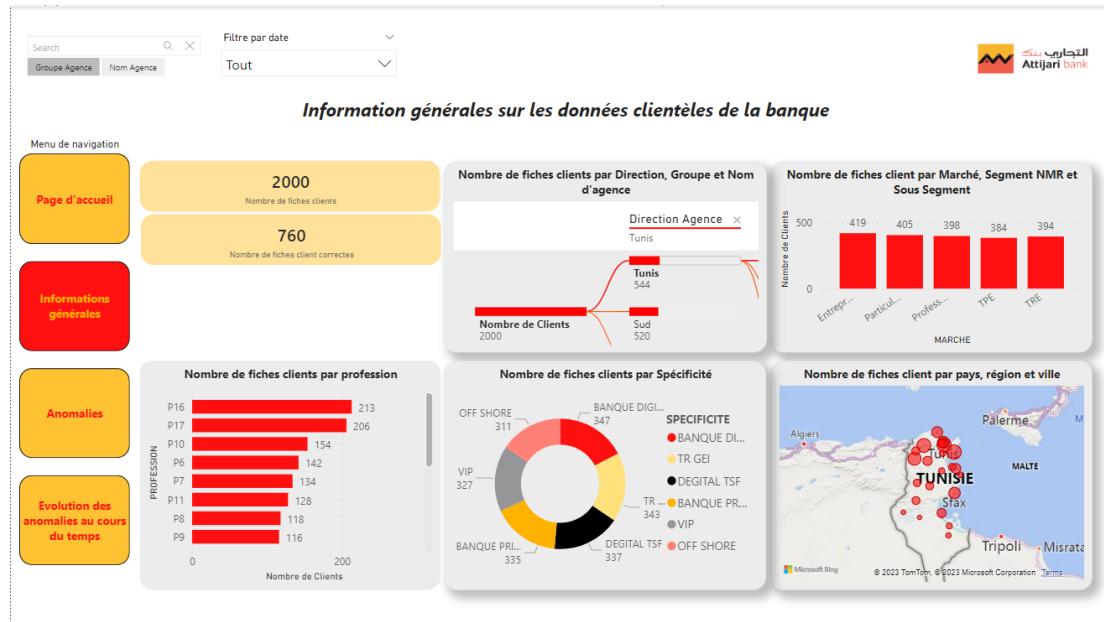


Figure 135 : Page « Information générales »

Le premier tableau de bord du rapport (figure 135) nous donne un aperçu général sur les fiches clientèles de Attijari Bank. Nous pouvons voir le nombre de fiche total ainsi que le nombre de fiches clients correctes (aucune anomalie).



Figure 136 : Répartition des fiches clients selon la direction régional, le groupe et le nom des agences

Le graphique (figure 136) illustre l'arborescence du nombre de fiches clients de la banque par agence. Il offre une vue détaillée de la répartition du nombre total de fiches clients selon trois niveaux hiérarchiques : la direction régionale, le groupe de l'agence et le nom spécifique de l'agence concernée.

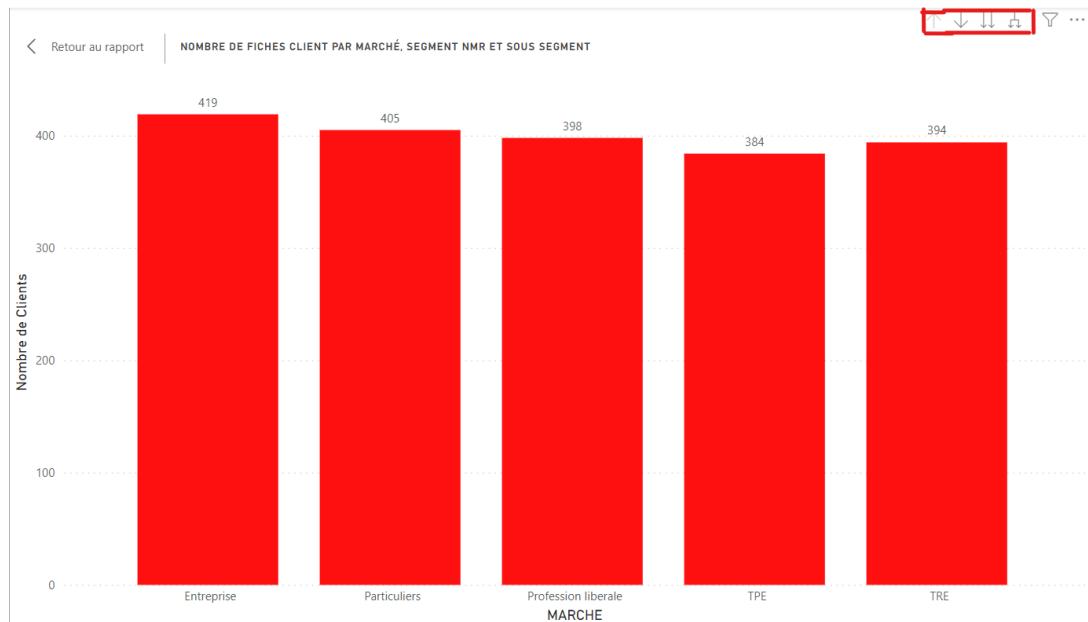


Figure 137 : Répartition des fiches clients selon le marché, le segment NMR et le sous segment

Le graphique (figure 137) est un histogramme qui illustre le nombre de fiches clients de la banque par marché, segment NMR et sous segment. Il offre une vue détaillée de la répartition du nombre total de fiches clients selon ces trois catégories. Nous pouvons descendre dans la hiérarchie pour accéder aux segments NMR et aux sous segments. L'interprétation possible à partir de ce graphe est que le marché d'entreprise présente le nombre le plus élevé de fiches clients. (Nous pouvons descendre dans la hiérarchie avec les boutons entourés en rouge).

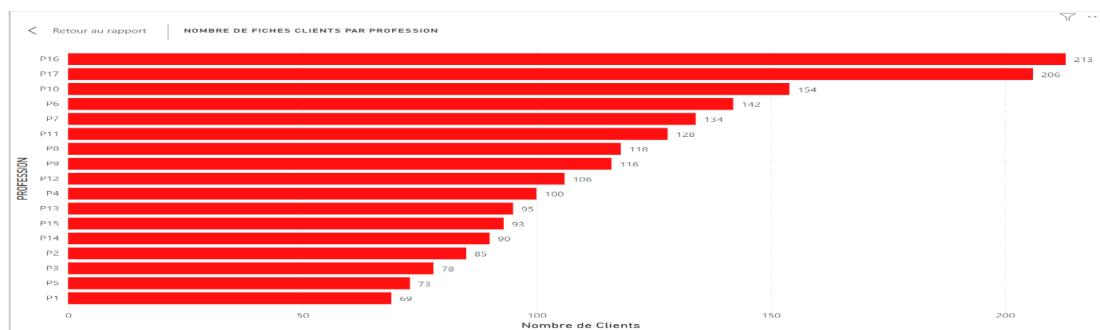


Figure 138 : Répartition des fiches clients selon leurs professions

Le graphique (figure 138) est un graphique à barres empilées qui illustre le nombre de fiches clients de la banque par profession. Il offre une vue détaillée de la répartition du nombre total de fiches clients selon leurs professions. L’interprétation possible à partir de ce graphe est que les clients dont la profession est « P16 » représentent le nombre le plus élevé de fiches client en termes de profession.

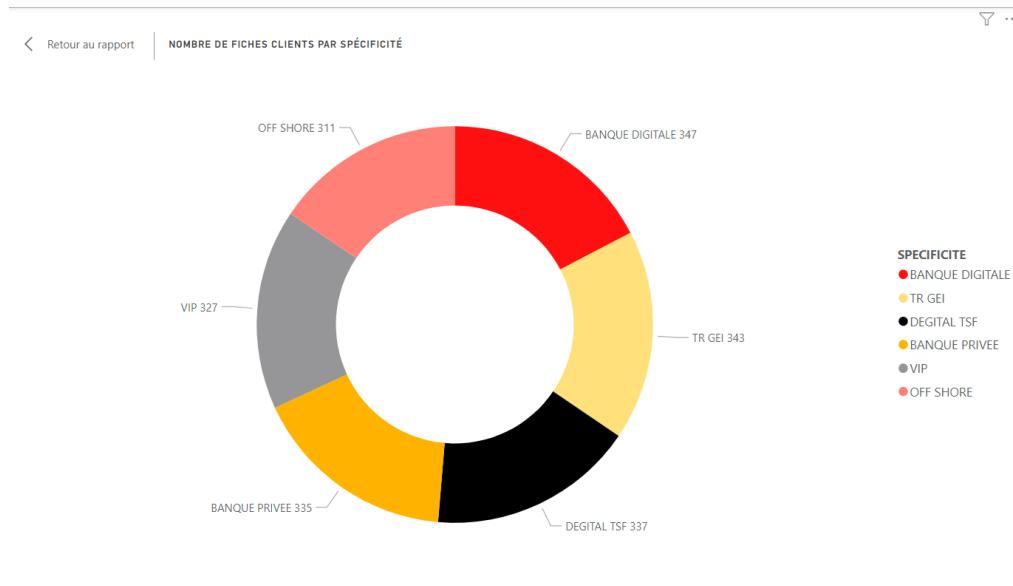


Figure 139 : Répartition des fiches clients selon leurs spécificités

Le graphique (figure 139) est un graphique en anneau qui illustre le nombre de fiches clients de la banque par spécificité. Il offre une vue détaillée de la répartition du nombre total de fiches clients selon leurs spécificités. L’interprétation possible à partir de ce graphe est que les clients dont la spécificité est « Banque Digitale » représentent le nombre le plus élevé de fiches client en termes de spécificité.

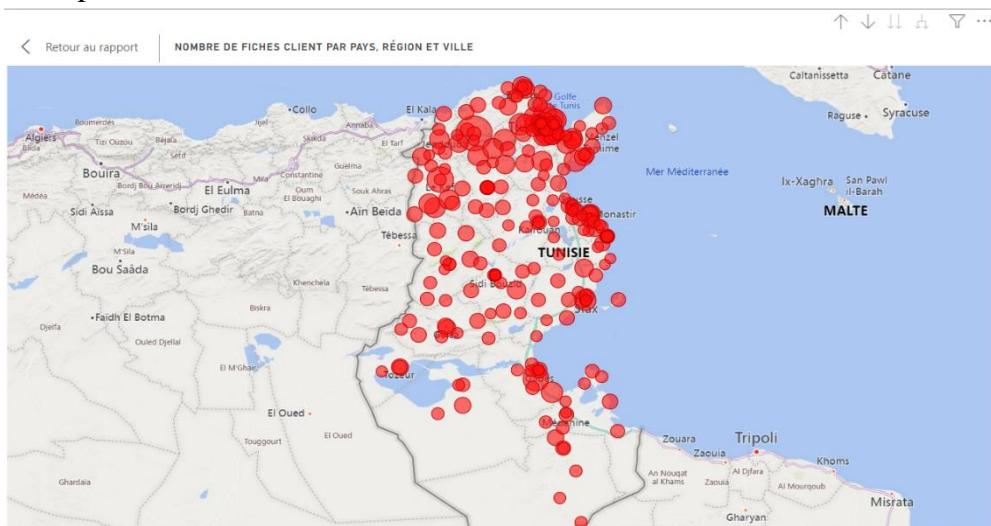


Figure 140 : Répartition des fiches clients selon leur pays, région et ville

Le graphique (figure 140) est une carte géographique qui illustre le nombre de fiches clients de la banque par pays, région et ville. Il offre une vue détaillée de la répartition du nombre total de fiches clients selon une hiérarchie de 3 niveaux : pays, région et ville. La taille des bulles est proportionnelle à la valeur du nombre de clients par pays, région et ville. Nous pouvons observer que la concentration des clients est plus élevé dans le Nord et le Cap Bon de la Tunisie.

Examinons maintenant les indicateurs qui composent la deuxième page de notre rapport. Cette page recense des données sur les anomalies des données clientèles par « use case », comme illustré dans la figure 141.

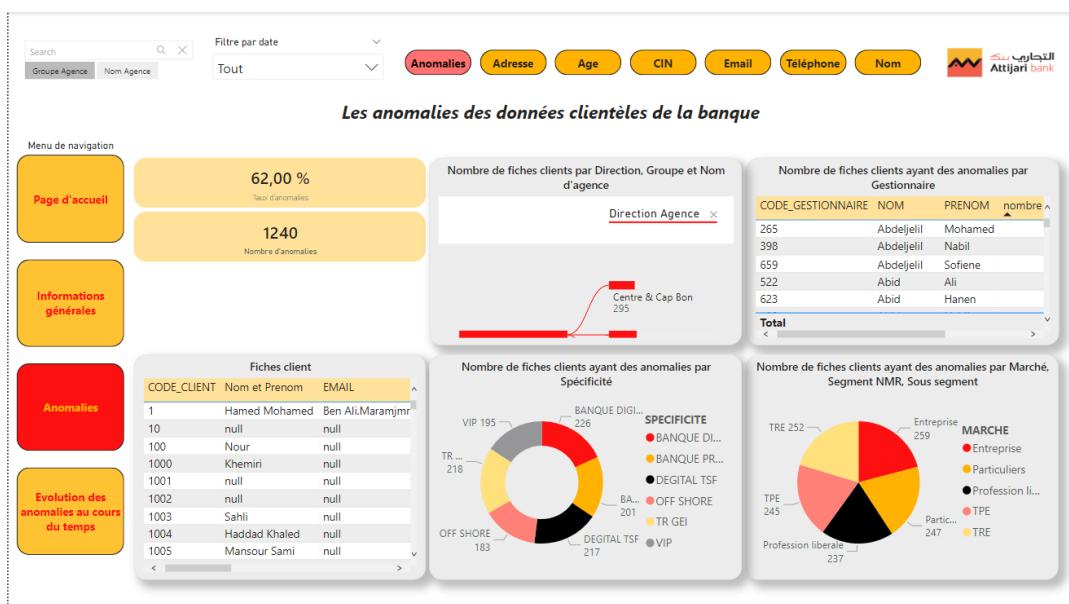


Figure 141 : Page « Anomalies »

Le deuxième tableau de bord du rapport (figure 141) nous donne un aperçu sur les fiches clientèles présentant des anomalies. Nous pouvons voir le nombre et le taux de fiches ayant des anomalies par rapport au nombre total des fiches.



Figure 142 : Répartition des fiches clients ayant des anomalies selon la direction régional, le groupe et le nom des agences

Le graphique (figure 142) illustre l'arborescence du nombre de fiches clients de la banque, ayant des anomalies, par agence. Il offre une vue détaillée de la répartition du nombre total de fiches clients selon trois niveaux hiérarchiques : la direction régionale, le groupe de l'agence et le nom spécifique de l'agence concernée. Nous pourrons à partir de ce graphe identifier la direction régionale, le groupe ou le nom de l'agence.

Retour au rapport | NOMBRE DE FICHES CLIENTS AYANT DES ANOMALIES PAR GESTIONNAIRE

CODE_GESTIONNAIRE	NOM	PRENOM	nombre de fiches client anomalies
356	Ben Ammar	Abdelhamid	7
499	Othmani	Riadh	7
942	Hammami	Khaled	6
368	Othmani	Sofiene	6
315	Sahli	Ons	6
327	Abdeljelil	Ahmed	5
450	Hajji	Khalil	5
737	Hajji	Riadh	5
831	Larbi	Yassmine	5
100	Mansouri	Yassmine	5
766	Toumi	Ons	5
958	Abdeljelil	Rania	4
622	Abid	Ons	4
817	Ben Ali	Wissem	4
190	Berriche	Nabil	4
402	Boussaid	Maram	4
645	Cherif	Adel	4
922	Cherif	Hanen	4
885	Cherni	Yassmine	4
478	Gharbi	Yosra	4
973	Haddad	Ameni	4
519	Haddad	Marwa	4
453	Haddad	Nabil	4
204	Haddad	Sofiene	4
570	Hajji	Abdelhamid	4
460	Hammami	Maissa	4
529	Khemiri	Marwa	4
787	Majri	Maissa	4
Total			1240

Figure 143 : Nombre de fiches clients ayant des anomalies par gestionnaires

Le graphique (figure 143) illustre un tableau qui présente le nombre de fiches clients ayant des anomalies par gestionnaires. Nous pourrons à partir de ce graphe identifier le gestionnaire ayant saisi le nombre le plus grand de fiches clients ayant des anomalies.

Retour au rapport | FICHES CLIENT

CODE_CLIENT	Nom et Prenom	EMAIL	NUM_CIN	ADRESSE	TELEPHONE_FIABLE	TELEPHONE_PORTABLE	Tel
1	Hamed Mohamed	Ben Ali.Maramjmriu2@hotmail.fr	null	18 janvier 6152 Rohia Siliana	null	null	1
10	null	null	null	Adissi 7170 Dahmani Le Kef	null	null	1
100	Nour	null	null	Akhout Mine 6111 Gaafour Siliana	null	null	1
1000	Khemiri	null	null	null	null	null	1
1001	null	null	null	null	null	null	1
1002	null	null	null	null	null	null	1
1003	Sahli	null	null	null	000464 4 4	2160 0 88 39 1 3	0
1004	Haddad Khaled	null	null	null	+21647 850 4 93	+2164 5 9 9 6 015	0
1005	Mansour Sami	null	null	null	null	null	1
1006	null	null	null	null	null	null	1
1007	null	null	null	null	null	null	1
1008	Wissem	null	null	null	null	null	1
1009	null	null	null	null	null	null	1
101	Yassmine	null	null	Akouda 4022 Akouda Sousse	null	null	1
1010	null	null	null	null	null	null	1
1011	Hajji Yosra	null	null	null	null	null	1
1012	null	null	88079302	null	null	null	1
1013	Boussaid Saber	Ben Ali.Mohamedp7pbmy@hotmail.fr	null	null	null	null	1
1014	null	null	null	null	41 8 0 91 8 4	+21671 42858 5	0
1015	null	null	null	null	null	null	1
1016	Haddad	null	null	null	null	null	1
1017	Mansouri	null	null	null	null	null	1
1018	null	null	null	null	null	null	1
1019	null	Ben Ali.Mariemmk2vg@yahoo.fr	null	null	null	null	1
102	Samir	null	null	Ali Bach-hamba 1046 Djebel Djelloud Tunis	0 00 438 6 5	02 74 415 6	0

Figure 144 : Fiches Clients

Le graphique (figure 144) illustre un tableau qui présente la liste des fiches clients. Nous pourrons à partir de ce graphe identifier les fiches clients ayant des anomalies et identifier les « use case » de ces anomalies (si un champ présent la valeur « null » c'est-à-dire qu'il n'est pas

une anomalie). Nous pouvons aussi savoir les fiches clients avec des anomalies associées à un gestionnaire en le sélectionnant dans le graphe qui précède celui-ci et la liste des fiches s'affiche dans ce graphe. Nous pouvons ensuite identifier les anomalies que présente la fiche du client.

Pour pouvoir naviguer entre différents tableau de bord dans la même page, nous avons utilisé la notion de Bookmark. Chaque bookmark représente une page et il est lié à un bouton.

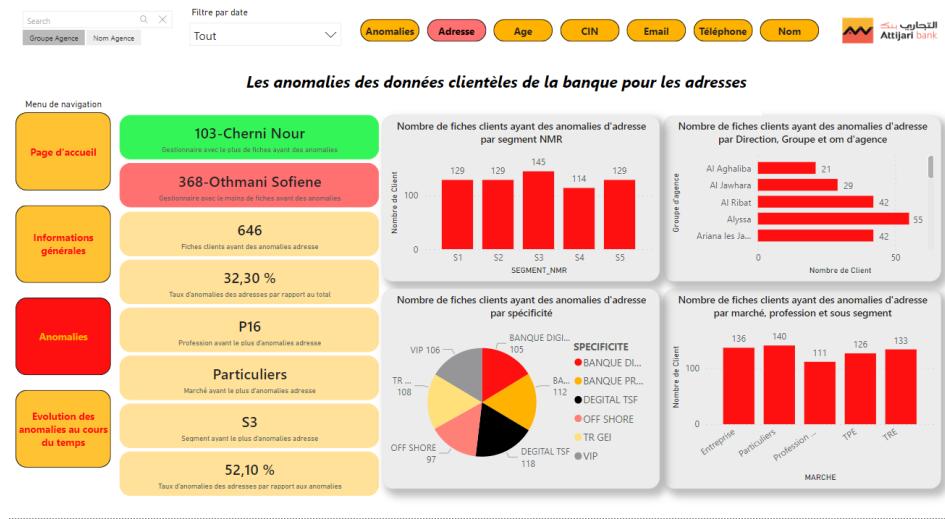


Figure 145 : Bookmark « Adresse »

Ce bookmark (figure 145) nous donne un aperçu sur les fiches clientèles présentant des anomalies dans leurs adresses. Nous pouvons voir le nombre et le taux de fiches ayant des anomalies dans les adresses des clients par rapport au nombre total des fiches. Nous avons aussi le taux de ces fiches par rapport au nombre total des anomalies.

Ce tableau de bord présente aussi la profession, le segment NMR et le marché ayant le plus de fiches avec anomalie adresse.

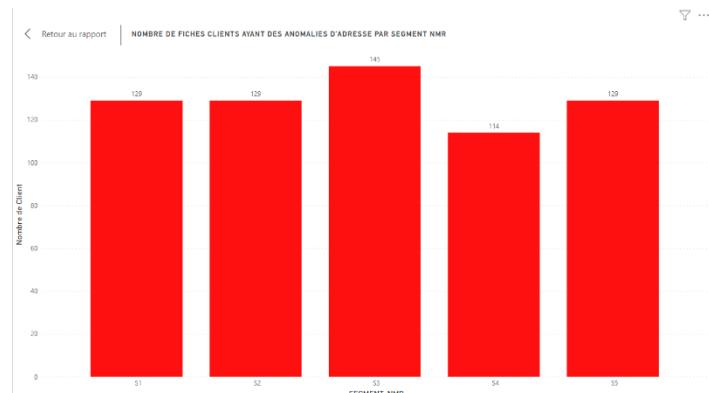


Figure 146 : Répartition des fiches clients ayant des anomalies dans leurs adresses selon le segment NMR

Le graphique (figure 146) est un histogramme qui illustre le nombre de fiches clients ayant des anomalies dans leurs adresses par segment NMR. L’interprétation possible à partir de ce graphe est que le segment NMR S3 présente le nombre le plus élevé de fiches clients ayant des anomalies dans leurs adresses.

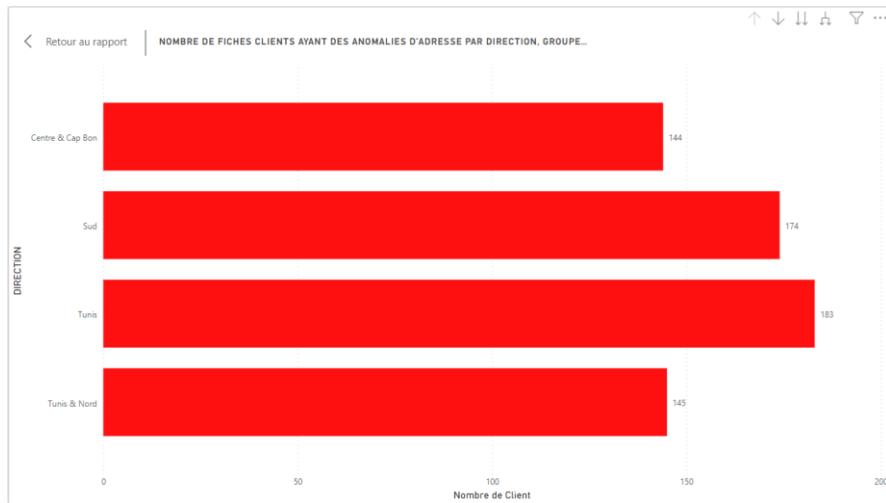


Figure 147 : Répartition des fiches clients ayant des anomalies dans leurs adresses selon la direction, le groupe et le nom d’agence

Le graphique (figure 147) est un graphique à barres empilées qui illustre le nombre de fiches clients ayant des anomalies dans leurs adresses de la banque par direction, groupe et nom d’agence. Nous pouvons descendre dans la hiérarchie pour accéder aux groupes et aux noms des agences. L’interprétation possible à partir de ce graphe est que les agences de Tunis présentent le nombre le plus élevé de fiches clients ayant des anomalies dans leurs adresses.

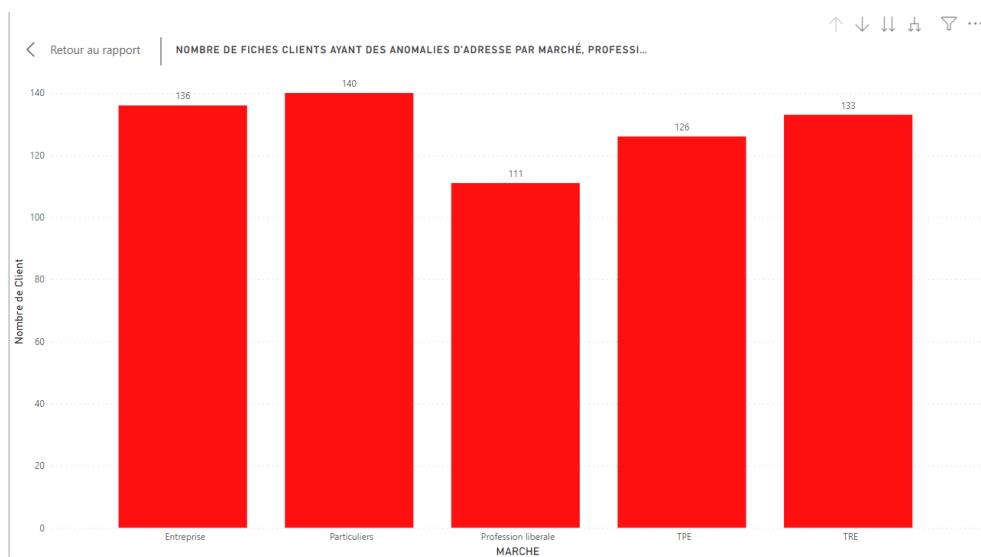


Figure 148 : Répartition des fiches clients ayant des anomalies dans leurs adresses selon le marché, la profession et le sous segment

Le graphique (figure 148) est un histogramme qui illustre le nombre de fiches clients, ayant des anomalies dans leurs adresses, par marché, profession et sous segment. Il offre une vue détaillée de la répartition du nombre total de fiches clients ayant des anomalies dans leurs adresses selon ces trois catégories. Nous pouvons descendre dans la hiérarchie pour accéder aux segments NMR et aux sous segments. L’interprétation possible à partir de ce graphe est que le marché des particuliers présente le nombre le plus élevé de fiches clients ayant des anomalies dans leurs adresses.

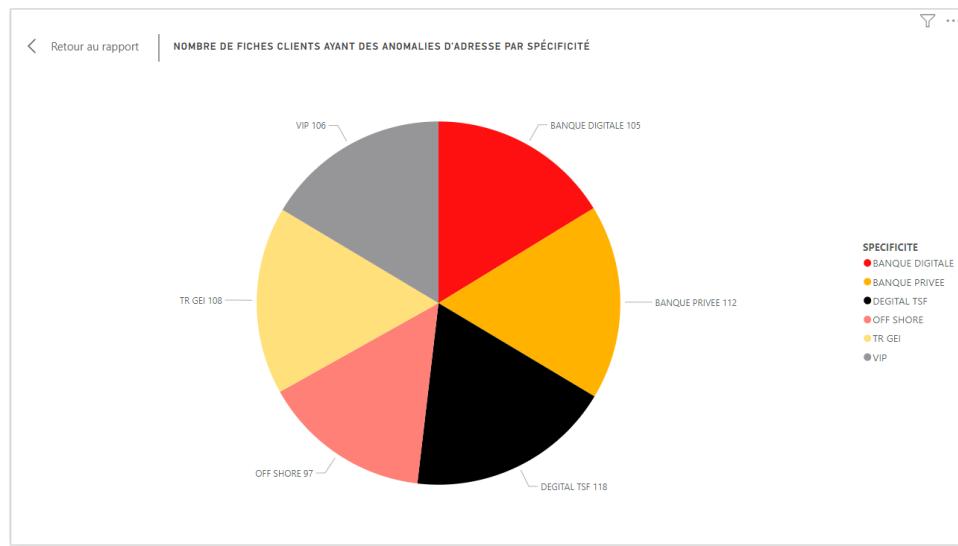


Figure 149 : Répartition des fiches clients ayant des anomalies dans leurs adresses selon leurs spécificités

Le graphique (figure 149) est un graphique en anneau qui illustre le nombre de fiches clients, ayant des anomalies dans leurs adresses, par spécificité.

Tous les visuels de ce bookmark sont filtrés selon l’adresse (comme nous avons déjà mentionné que les adresses présentant une chaîne « null » sont en réalité des valeurs correctes par conséquent nous pouvons filtrer les adresses dont les valeurs ne sont pas une chaîne « null » pour avoir les adresses incorrectes) (figure 150).



Figure 150 : Filtre des adresses

Les mêmes visuels seront présents dans chaque bookmark tout en changeant le filtre à chaque fois selon le type d'anomalie. Les bookmarks sont présentés dans les figures de 151 à 155.

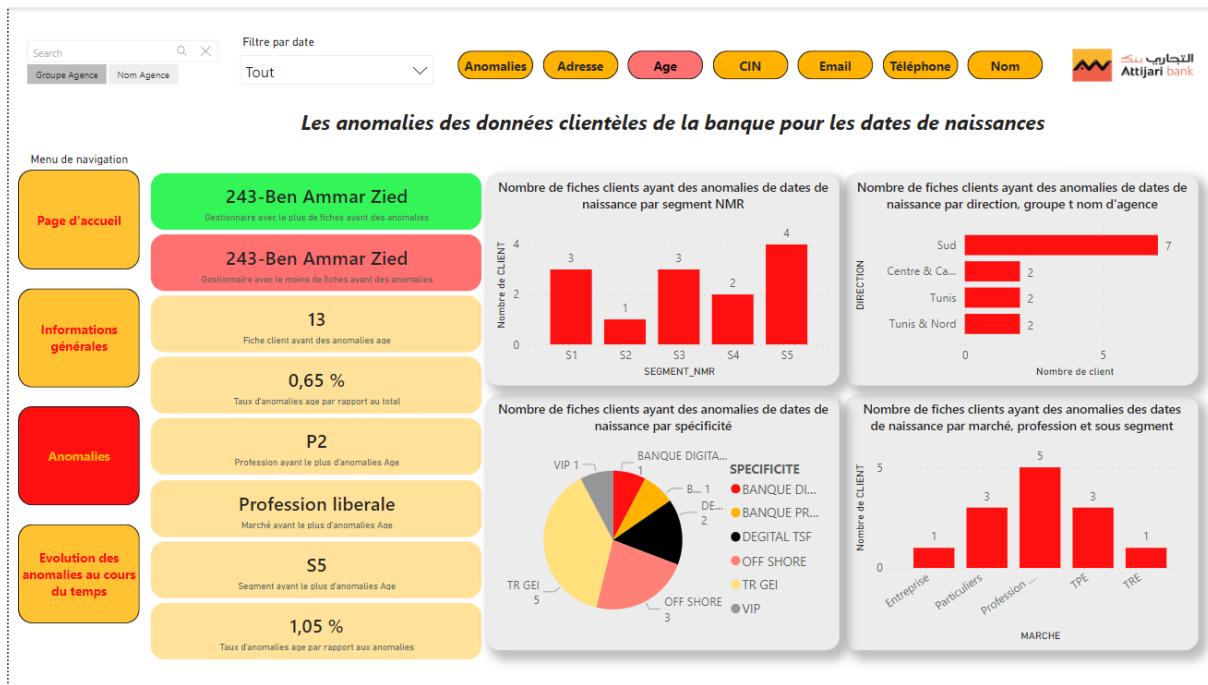


Figure 151 : Bookmark « Date de naissance »

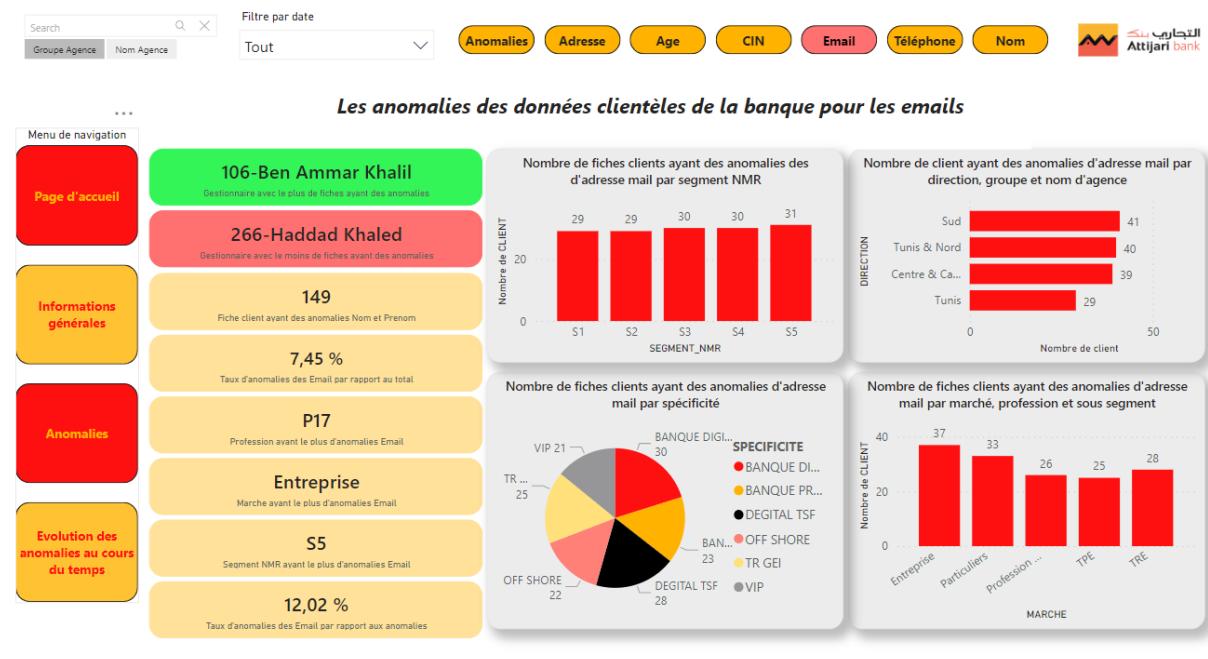


Figure 152 : Bookmark « Email »

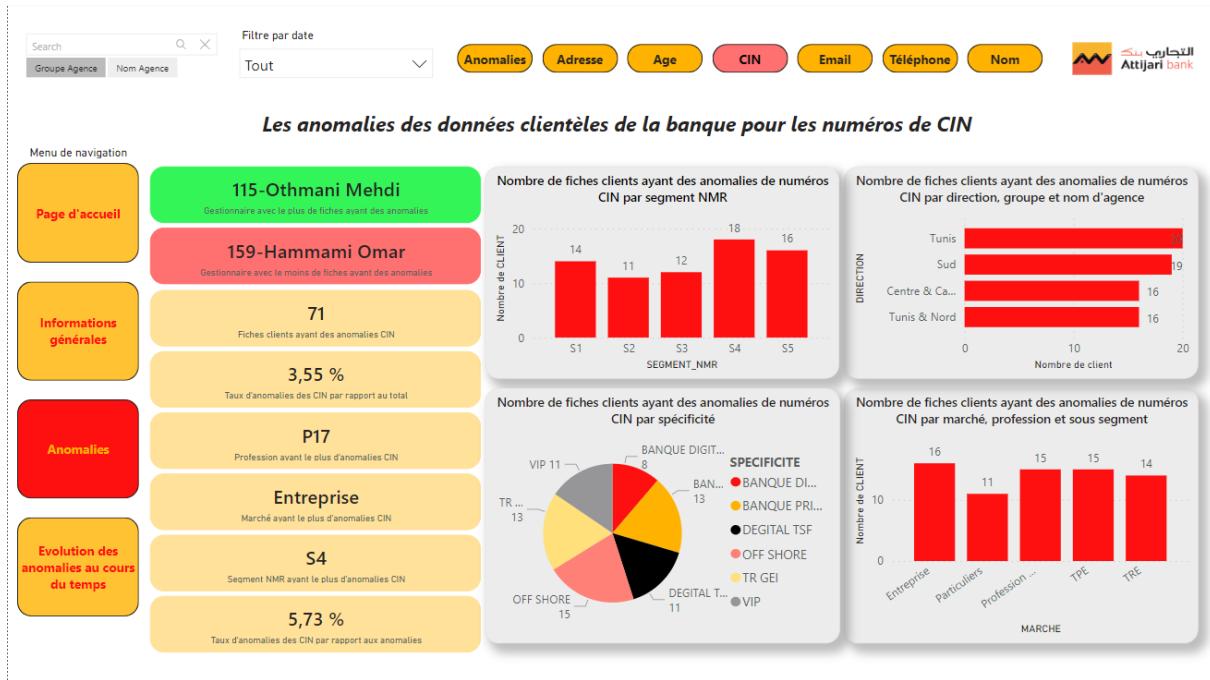


Figure 153 : Bookmark « CIN »

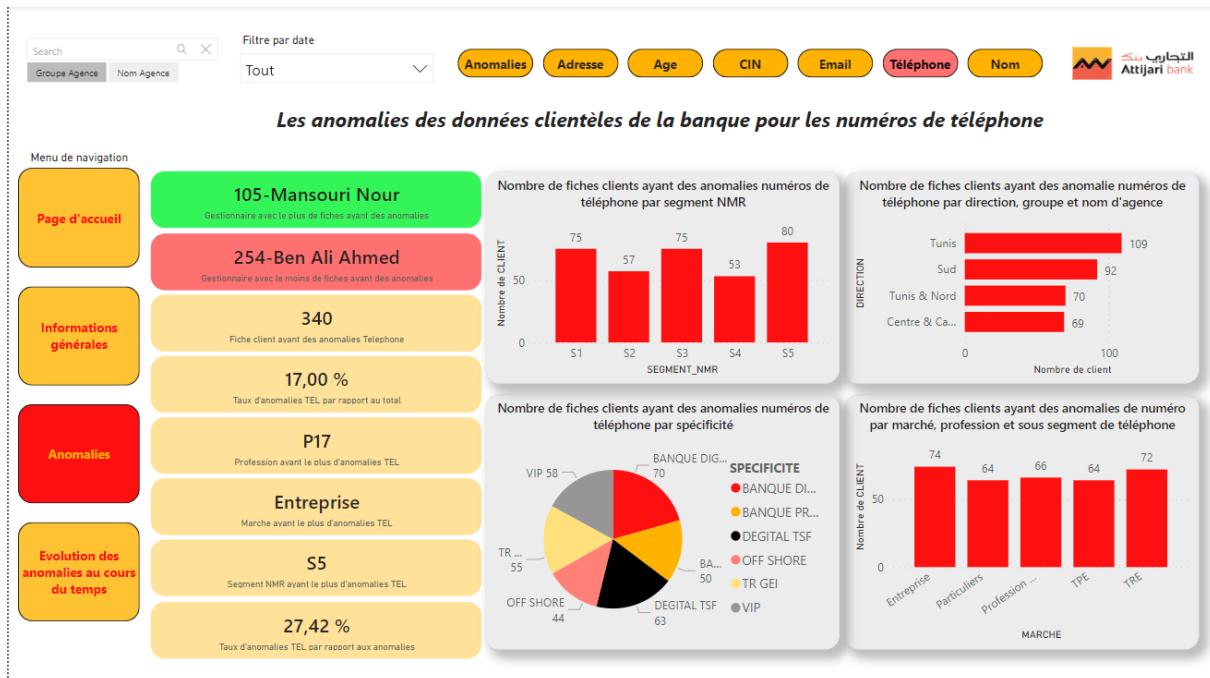


Figure 154 : Bookmark « Téléphone »

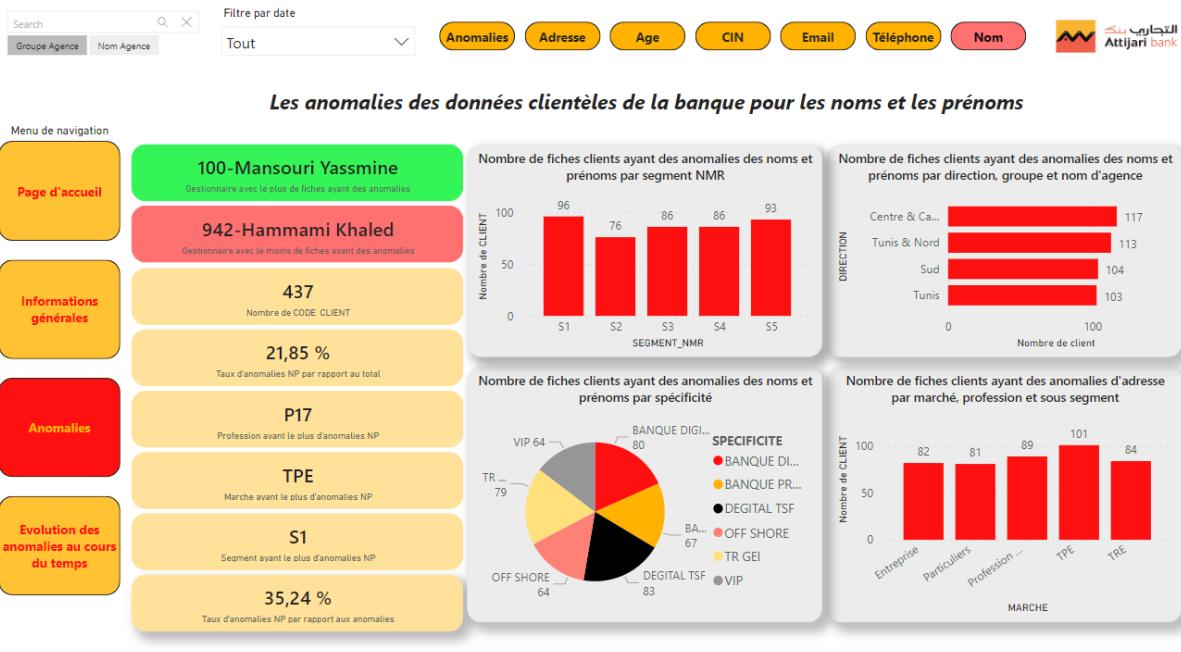


Figure 155 : Bookmark « Nom et prénom »

Examinons maintenant les indicateurs qui composent la troisième page de notre rapport. Cette page recense des données sur l'évolution temporelle des anomalies des données clientèles par type d'anomalie comme illustré dans la figure 156.

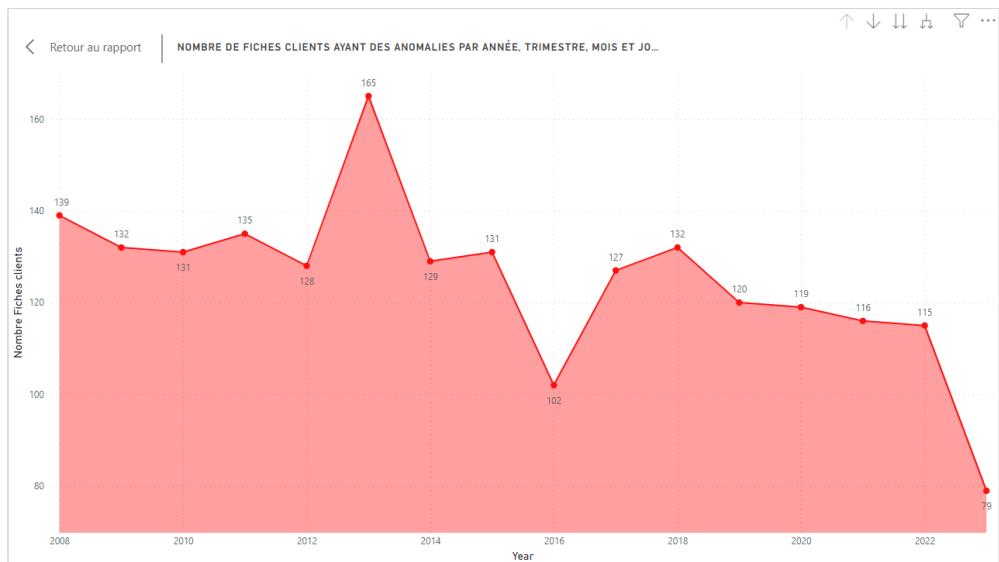


Figure 156 : Page « Evolution du nombre des anomalies au cours du temps »

Le troisième tableau de bord du rapport (figure 156) nous donne un aperçu sur l'évolution du nombre des anomalies au cours du temps.



Le graphique (figure 157) est un graphique en aires qui illustre le nombre de fiches clients ayant des anomalies selon la date de leurs modifications par année, trimestre, mois et jour. Nous pouvons descendre dans la hiérarchie pour accéder aux niveaux plus bas. L'interprétation possible à partir de ce graphe est qu'en 2013 nous avons le nombre le plus élevé de fiche clientèle présentant des anomalies (165 fiches avec anomalies). En revanche, 2023 présente le nombre le plus faible avec 79 fiches avec anomalies qui ont été modifiées .

Maintenant pour ce qui concerne les types d'anomalies, nous avons utilisé la même technique (celle des bookmarks). Les bookmarks sont présentés dans les figures de 158 à 163.

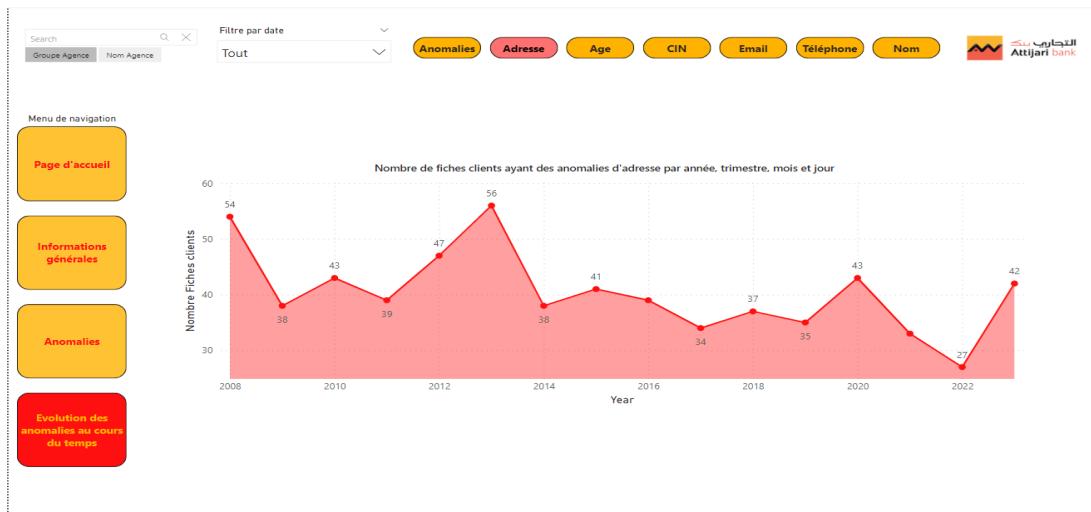




Figure 159 : Bookmark « Evolution du nombre des anomalies de date de naissance au cours du temps »

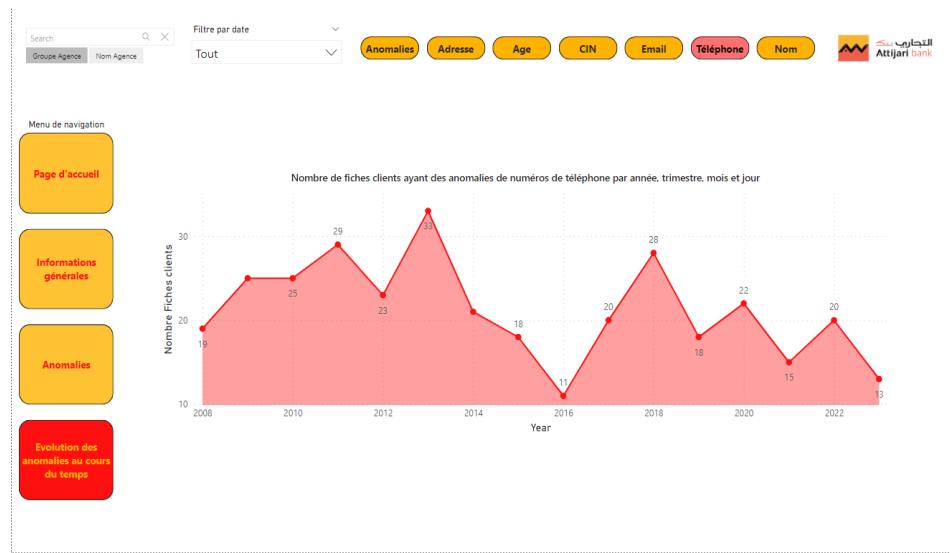


Figure 160 : Bookmark « Evolution du nombre des anomalies dans les numéros de téléphones au cours du temps »



Figure 161 : Bookmark « Evolution du nombre des anomalies dans les numéros de CIN au cours du temps »



Figure 162 : Bookmark « Evolution du nombre des anomalies dans les nom et prénoms au cours du temps »



Figure 163 : Bookmark « Evolution du nombre des anomalies email au cours du temps »

Finalement, Nous avons opté à l'utilisation de deux filtres (figure 164). Un premier filtre pour les groupes et noms des agence et un deuxième filtre pour les dates (par année, trimestre, mois et jour). Nous pouvons appliquer ces filtres sur toutes les pages.

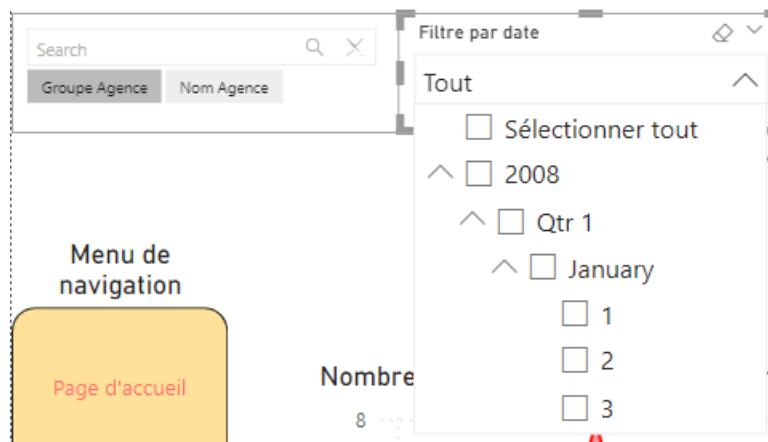


Figure 164 :Filtres agence et date

Conclusion

Ce quatrième chapitre nous a permis de donner aux décideurs une vue globale sur la répartition des anomalies des données clientèle dans la base de données de la banque, en utilisant l'outil de reporting Power BI, pour pouvoir détecter la plus grande concentration des données ayant des anomalies afin d'améliorer la qualité de données.

Conclusion générale

Pour conclure ce rapport, nous avons réalisé ce projet de fin d'études pour l'obtention d'une licence en Business Intelligence à Attijari Bank. Nous avions comme mission principale de créer un système décisionnel pour la détection des anomalies des données clientèle de la banque. Ça nous a pris quatre mois pour concevoir ce projet.

Tout au long du premier chapitre, nous avons procédé à l'identification du contexte du projet. Dans le but de bien gérer ce projet et concevoir une solution décisionnelle digne de l'entreprise d'accueil. On a pris le temps de comprendre les besoins de la banque, l'étude de l'existant ainsi que l'output final souhaité par les décideurs. Ensuite on a pris comme décision de suivre une méthodologie de gestion de projet adaptée à la nature du travail demandé. On a donc mené ce projet selon la méthode GIMSI.

Le deuxième chapitre de ce rapport s'intitule Conception de la solution. Tout au long de ce chapitre nous avons présenté la conception de notre solution. Nous avons commencé par la définition de certaines notions théoriques de la BI. Ensuite nous avons parlé des indicateurs et nous avons identifié ceux que nous allons utiliser dans les visuels. Par la suite nous avons présenté le modèle de notre Data Warehouse. Et finalement nous avons mis le prototype des tableaux de bord.

Lors du troisième chapitre nous avons travaillé sur l'alimentation de notre Data Warehouse. Nous avons commencé par la simulation des données sources. Ensuite nous avons entamé le processus ETL par l'extraction des données à partir des fichiers sources, la transformation des données en appliquant des règles de gestion et finalement le chargement des données dans le Data Warehouse.

Finalement nous avons travaillé sur le dernier chapitre intitulé Restitution des données. Nous avons présenté les différents tableaux de bord de notre rapport Power BI.

Nous estimons que cette solution aura une valeur ajoutée pour les analystes et les décideurs de Attijari Bank vu son implémentation avec les données réelles de la banque, mais par contrainte de confidentialité nous ne pouvons pas afficher les résultats.

Cette solution peut être considérée comme un premier pas dans un projet de « Data Quality », nous pouvons ensuite mettre en place un modèle de Data Mining qui permet par exemple d'identifier si un numéro de CIN (carte d'identité nationale) est fiable ou non en fonction d'autres critères comme la date de naissance du client ou bien la date de délivrance de la carte, ce qui nous permet d'identifier encore plus les anomalies dans les numéros de CIN.

Je tiens à remercier les membres du jury pour leur présence et leur lecture attentive de ce rapport. Ce stage était une opportunité pour découvrir le monde professionnel et aussi de mettre en pratique les compétences que j'ai acquises durant mon processus universitaire.

Webographie

[1] Définition et présentation de la méthodologie GIMSI :

<https://www.piloter.org/mesurer/methode/methode-business-intelligence.htm> → visité le 02/03/2023

<https://www.piloter.org/livres-blancs-pdf/methode-business-intelligence.pdf> → visité le 02/03/2023

[2] Définition et présentation de la méthodologie SCRUM :

<https://bubbleplan.net/blog/agile-scrum-gestion-projet/> → visité le 04/03/2023

[3] Comparatif des outils ETL :

<https://www.cartelis.com/blog/comparatif-logiciels-etl/> → visité le 16/03/2023

[4] Comparatif des outils de visualisation des données :

[Comparatif Qlik Sense vs Power BI \(next-decision.fr\)](https://www.next-decision.fr/comparatif-qlik-sense-vs-power-bi/) → visité le 07/04/2023

[5] Bonnes pratiques sur Talend :

<https://www.next-decision.fr/wiki/bonnes-pratiques-talend> → visité le 19/03/2023

[6] Caractéristiques d'un indicateur :

[Comment définir de bons indicateurs pour vos projets ? - Well Grounded \(well-grounded.org\)](https://www.well-grounded.org/article/comment-definir-bons-indicateurs-pour-vos-projets) → visité le 27/03/2023

Résumé

Le présent projet de fin d'étude intitulé « conception et développement d'une solution bi pour la détection des anomalies des données clientèle de la banque » a pour vocation de donner aux décideurs une vision globale sur la répartition des anomalies clientèles de la banque, afin de pouvoir améliorer la qualité de données au sein de la banque.

Nous avons commencé par la simulation des données fictives vu la contrainte de la confidentialité des données de la banque. Ensuite nous avons appliqué le processus ETL par l'extraction des données des fichiers sources, l'application des règles de gestion pour transformer les données et le chargement de ces données dans l'entrepôt de données. On a utilisé le logiciel Talend Open Studio for Data Integration pour appliquer ce processus.

Et pour finir nous avons présenté la répartition des anomalies des données clientèles dans un rapport composé de plusieurs tableaux de bord pour avoir une vision globale qui aide à la prise de décision.

Mots-clés: ETL, Talend, Power BI, qualité de données, règles de gestion, anomalies, tableaux de bord, entrepôt de données.

Abstract

The aim of this end-of-study project, entitled "Design and development of a bi solution for detecting anomalies in the bank's customer data", is to give decision-makers a global view of the distribution of customer anomalies in the bank, so that they can improve data quality within the bank.

We began by simulating fictitious data, given the bank's data confidentiality constraints. Then we applied the ETL process, extracting data from the source files, transform the data and loading it into the data warehouse. Talend Open Studio for Data Integration was used to apply this process.

Finally, we presented the distribution of customer data anomalies in a report composed of several dashboards, to provide a global view that aids decision-making.

Keywords: ETL, Talend, Power BI, data quality, anomalies, dashboards, data warehouse.