

**Entreprise : SIMPLON.CO**

**Direction : Système d'information (SI)**

**Période de stage : du 09 juin 2025 au 10 décembre 2025**

## Formez-vous à la Tech simplement

Simplon vous accompagne simplement pour développer vos compétences numériques et rend les métiers de la tech accessibles à toutes et à tous.

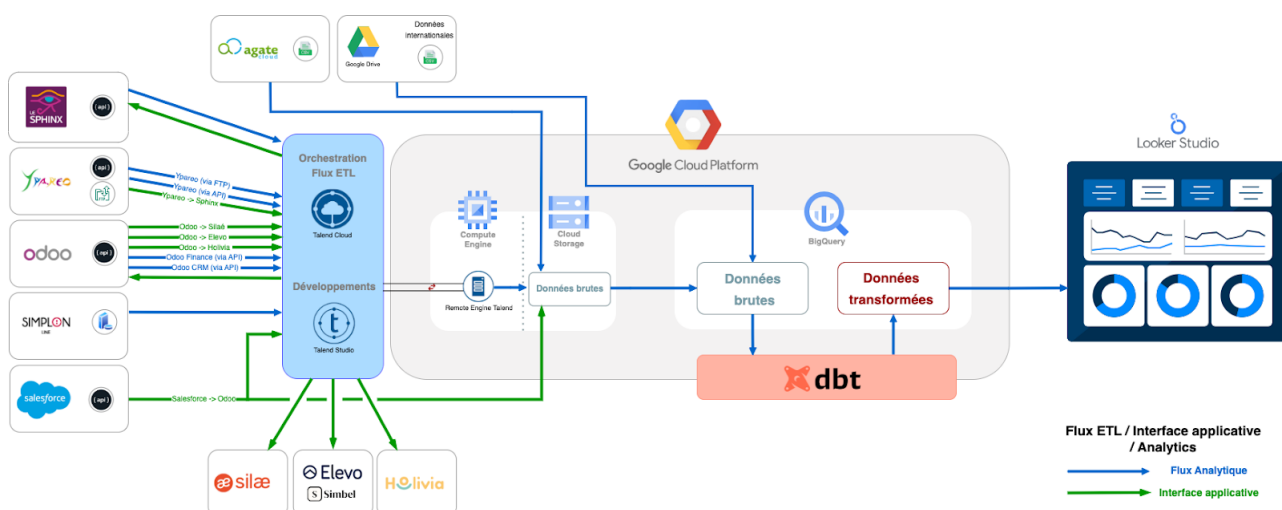
Du niveau Découverte au niveau Bac +5, rejoignez nos formations en Développement, Réseaux & Cybersécurité, Data / IA.

[Candidater](#)

[Découvrir les filières](#)



## Stage Data – Gestion, modélisation de données et Datavisualisation



**Rapport rédigé par**  
Ahmed HADAD

**Tutrice en entreprise**  
Juliana CAVALCANTE

**Tutrice universitaire**  
Sophie DABO

## Remerciements

Je tiens à exprimer ma profonde gratitude envers toutes les personnes qui m'ont accompagné tout au long de mon stage au sein de la **direction du Système d'Information (SI) de Simplon.co**.

Je remercie tout particulièrement ma tutrice, **Juliana CAVALCANTE**, pour ses conseils avisés, son encadrement attentif et sa disponibilité constante, ainsi que **Antony HENRION** pour son accompagnement technique et ses précieux retours sur mes travaux.

Je souhaite également remercier **Hanifa ABDELAZIZ**, directrice de la **direction du Système d'Information (SI) de Simplon.co**, pour son soutien et sa bienveillance de l'équipe SI, qui ont contribué à la réussite de ce stage et à mon intégration au sein de l'équipe.

Ce stage m'a permis de développer mes compétences en **gestion, contrôle et modélisation de données** via *DBT* et *Google Cloud Platform*, ainsi qu'en visualisation de données avec *Looker Studio*. Il m'a également offert un contact permanent avec les équipes métiers, ce qui m'a permis de répondre efficacement à leurs besoins de ticketing.

Ce stage m'a permis donc de couvrir l'ensemble du cycle de la donnée : de la collecte et du contrôle de qualité à l'analyse, la modélisation et la restitution visuelle, en développant rigueur, précision. Il m'a surtout permis de comprendre les fonctionnalités des outils techniques du métiers, notamment : ***Google Cloud Platform*** pour le stockage, le maintien et l'analyse des données massives, ***DBT*** pour la modélisations de données et ***Looker Studio*** pour la visualisation des données.

Les échanges avec les équipes métiers et la participation aux ateliers de data visualisation m'ont offert une expérience concrète et enrichissante, renforçant ma compréhension des enjeux de la donnée en entreprise.

Enfin, je remercie l'ensemble de la direction du Système d'Information (SI) de Simplon.co pour leur accueil chaleureux et leur collaboration tout au long de cette expérience professionnelle.

# INTRODUCTION

Dans le cadre de mon stage au sein de l'équipe **Data** de la direction du Système d'Information (SI) de Simplon.co, j'ai eu l'occasion de travailler sur plusieurs projets visant à améliorer la qualité, la cohérence, la modélisation et la visualisation des données utilisées par les équipes opérationnelles, pédagogiques, administratives et commerciales.

Comme dans toute organisation, la gestion et l'exploitation de la donnée constituent des enjeux majeurs pour **Simplon.co** : pilotage des formations, suivi des candidat · es, optimisation des placements ou encore diffusion d'outils internes. Mon rôle a principalement consisté à contribuer à la modélisation des données, à l'automatisation de leur traitement ainsi qu'à la création de tableaux de bord à destination des équipes métiers.

Je présenterai dans ce rapport certaines missions réalisées au cours de mon stage, notamment un projet central intitulé **Projet PLACEMENT**, ainsi que plusieurs **travaux transverses**.

Ce rapport expose les enjeux, les approches techniques mobilisées et les résultats obtenus pour chacun de ces travaux. Il est structuré comme suit :

1. Une brève présentation de **Simplon.co** et du contexte métier,
2. Trois chapitres composés de :
  - (a) **Cycle de la donnée et environnement technique** : présentation des outils techniques utilisés et du cycle d'étude au sein de l'équipe Data.
  - (b) **Projet PLACEMENT** : mise en place de tableaux de bord interactifs permettant notamment :
    - a) aux commerciaux d'identifier rapidement les profils des candidat · es en recherche d'entreprise (filtres par région, formation, etc.), adaptés dès qu'un besoin est détecté chez un client. Le tableau permet également de suivre les sessions ouvertes en alternance ;
    - b) au directeur du pôle alternance à distance de suivre les statistiques de placement des sessions en alternance.
  - (c) **Travaux transverses** :
    - a) résolution de tickets : réponses aux besoins des équipes (exports, analyses, vérifications spécifiques) ;
    - b) mise en place de contrôles de données permettant d'assurer la fiabilité des indicateurs.
3. Une conclusion

# Présentation de Simplon.co

## 1. Présentation de l'entreprise

Créée en 2013, **Simplon.co** est un réseau de fabriques numériques engagé pour un numérique inclusif. L'organisation forme aux métiers du numérique, notamment des publics éloignés de l'emploi, via des bootcamps intensifs fondés sur une pédagogie active et par projets. Aujourd'hui, Simplon.co constitue l'un des plus grands réseaux de bootcamps inclusifs au monde, présent dans **34 pays** avec plus de **84 écoles** et plus de **19 937 apprenant · es** formé · es à l'international.

## 2. Organisation internationale

L'activité internationale de Simplon.co s'appuie sur trois hubs, dont deux principaux :

- a) **Maghreb Coding School** (Maroc), opérant des programmes sur l'ensemble du territoire ;
- b) **Simplon Africa**, implantée dans plusieurs pays d'Afrique subsaharienne.

Ces hubs pilotent des projets de formation adaptés aux besoins locaux, en partenariat avec des acteurs publics, privés ou institutionnels.

## 3. Projets et partenariats clés

Simplon.co mène une large variété de projets : consortiums européens, programmes financés par l'**AFD** ou l'**Union européenne**, déploiement de bootcamps "**Powered by Simplon**". Un exemple notable est le partenariat avec les **Orange Coding Academies** en Jordanie, où plus de **742 apprenant · es** ont été formé · es depuis 2019, avec un taux de **sortie positive** de **60%**.

## 4. Savoirs-faire et expertise

L'entreprise intervient sur l'ensemble de la chaîne de valeur d'un projet de formation numérique :

- a) études de besoins et faisabilité ;
- b) ingénierie pédagogique et création de parcours ;
- c) formation des formateurs ;
- d) dispositifs qualité, sourcing inclusif et certification ;
- e) mobilisation d'outils numériques, dont la plateforme pédagogique *Simplonline*.

Simplon.co combine ainsi expertise pédagogique, innovation sociale et outils numériques pour former aux compétences du futur et accompagner les transformations du marché de l'emploi.

## Table des matières

<b>1</b>	<b>Cycle de la donnée et environnement technique</b>	<b>6</b>
1.1	Cycle de la donnée . . . . .	6
1.1.1	Gestion des données . . . . .	6
1.1.2	Extraction et stockage des données . . . . .	6
1.2	Outils de modélisation, d'analyse et de visualisation . . . . .	7
1.2.1	Outils d'analyse et de reporting . . . . .	8
1.3	Stratégie de l'équipe DATA . . . . .	10
<b>2</b>	<b>Projet PLACEMENT</b>	<b>11</b>
<b>3</b>	<b>Travaux transverses</b>	<b>12</b>
3.1	Résolution des tickets . . . . .	12
3.1.1	Données certification Agilité . . . . .	12
3.1.2	Apple Reporting . . . . .	14
3.2	Contrôle des données internationales . . . . .	15
3.2.1	Données renseignées dans les fichiers pays . . . . .	15
3.2.2	Problèmes observés dans le tableau de bord existant . . . . .	16
3.2.3	Initiative de l'équipe Data . . . . .	16
<b>4</b>	<b>Conclusion</b>	<b>19</b>

## Cycle de la donnée et environnement technique

Avant de présenter les missions réalisées, il est essentiel de comprendre le cycle de vie de la donnée chez Simplon.co : d'où proviennent les données ? Quels outils techniques sont mobilisés ? Comment sont-elles migrées et exploitées par l'équipe DATA ?

### 1.1 Cycle de la donnée

Chez Simplon.co, le cycle de la donnée permet de collecter, de contrôler, de modéliser et de visualiser la donnée afin d'analyser les compétences, les évaluations, les apprenants, les formations et d'accompagner les décisions stratégiques de manière fiable :

#### 1.1.1 Gestion des données

Les données sont hébergées dans différents outils internes : Odoo, CRM, Yparéo, Sphinx et Simplonline. Chacun possède un rôle spécifique selon le type de données stockées.

1. **Sphinx** :  
Plateforme permettant de collecter des données issues de questionnaires. Simplon suit attentivement ses apprenants après la formation : satisfaction, suggestions, insertion professionnelle, etc.
2. **Yparéo** :  
Base de données contenant les informations sur les apprenants, les groupes et les sessions de formation.
3. **Odoo** :  
Fournit des données concernant les employés (dont les formateurs), les finances et le CRM (inscriptions).
4. **Simplonline** :  
Plateforme pédagogique permettant de suivre l'activité des apprenants.

#### 1.1.2 Extraction et stockage des données

Les données sont extraites via l'outil ETL *Talend*. Les flux sont mis à jour quotidiennement.

**Note** : Cette étape est du ressort du Data Engineer. Dans ce volet, mon rôle consistait à contrôler et à alerter en cas d'échec ou d'incohérence détectée.

Les données sont ensuite déposées telles quelles dans *Google Cloud Storage (GCS)*, sans traitement préalable.

## 1.2 Outils de modélisation, d'analyse et de visualisation

Une fois déposées dans *Google Cloud Storage (GCS)*, les données brutes sont utilisées pour créer des modèles structurés et optimisés répondant aux besoins des équipes métiers. Ces modèles sont développés dans *DBT* et hébergés sur *BigQuery*.

### DBT et GCP : Modélisation et hébergement de données

Ma mission principale consistait à transformer et structurer les données via *DBT* sur *BigQuery*. J'ai construit des modèles analytiques adaptés aux besoins métiers, reposant notamment sur des modèles à plat et en étoile.

#### 1. Modélisation à plat :

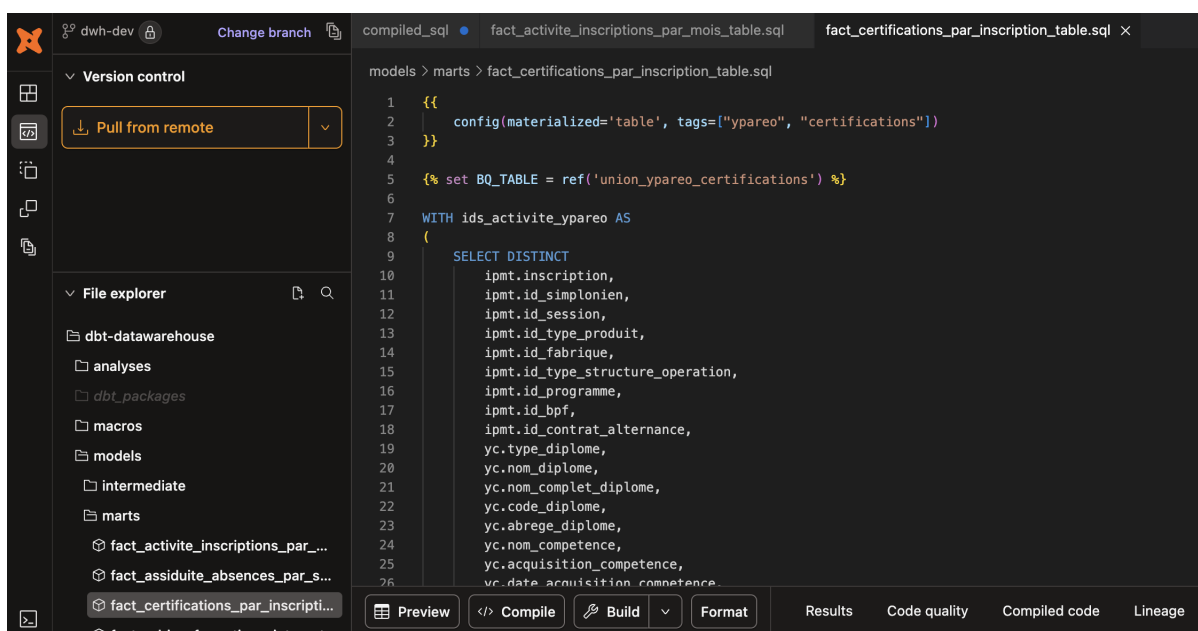
La modélisation à plat consiste à stocker toutes les informations dans une seule table, sans séparation entre faits et dimensions. Elle simplifie l'accès aux données mais entraîne souvent des redondances. Ce modèle est adapté aux analyses rapides ou aux datasets de petite taille.

#### 2. Modélisation en étoile :

La modélisation en étoile organise les données autour d'une table de faits centrale reliée à plusieurs tables de dimensions. Elle réduit la redondance, améliore les performances analytiques et facilite la compréhension des données. C'est le modèle le plus utilisé dans notre équipe DATA. Qu'est ce qu'une table de fait et de dimension ?

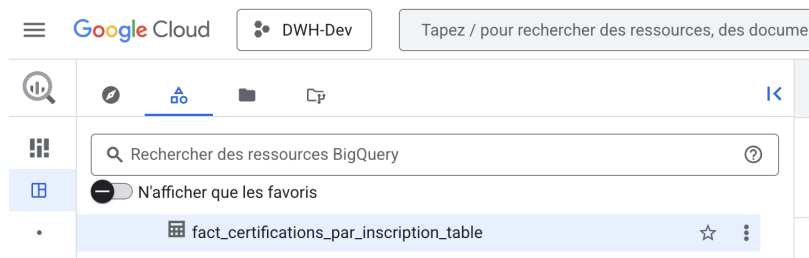
- (a) La table de faits contient les événements mesurables, comme chaque certification obtenue ou chaque inscription à une formation. Elle enregistre les résultats, dates et indicateurs de performance pour chaque apprenant. Les clés étrangères permettent de la relier aux dimensions (apprenant, session, formation).
- (b) Les tables de dimensions décrivent les entités de Simplon.co : les apprenants (nom, date de naissance, genre), les formations (titre, code programme), ou les sessions (dates, lieux). Elles servent à qualifier et filtrer les faits lors des analyses. Chaque dimension possède une clé primaire qui se relie aux faits.

Voici par exemple un modèle visant à structurer les informations d'un apprenant en lien avec la certification associée à sa formation :



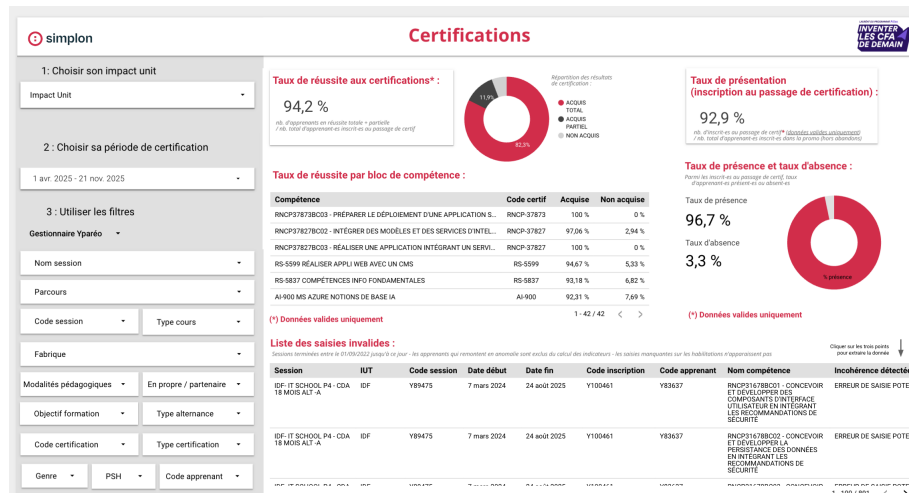
```
1  {{
2    config(materialized='table', tags=["ypareo", "certifications"])
3  }}
4
5  {% set BQ_TABLE = ref('union_ypareo_certifications') %}
6
7  WITH ids_activite_ypareo AS
8  (
9    SELECT DISTINCT
10     ipmt.inscription,
11     ipmt.id_simplonien,
12     ipmt.id_session,
13     ipmt.id_type_produit,
14     ipmt.id_fabrique,
15     ipmt.id_type_structure_operation,
16     ipmt.id_programme,
17     ipmt.id_bpf,
18     ipmt.id_contrat_alternance,
19     yc.type_diplome,
20     yc.nom_diplome,
21     yc.nom_complet_diplome,
22     yc.code_diplome,
23     yc.abrege_diplome,
24     yc.nom_competence,
25     yc.acquisition_competence,
26     yc.date_acquisition_competence
```

Le modèle finalisé est ensuite hébergé dans *BigQuery*, prêt à être interrogé :



## Looker Studio : Visualisation

Les modèles sont utilisés dans *Looker Studio* afin de produire les tableaux de bord nécessaires au pilotage des indicateurs métiers.



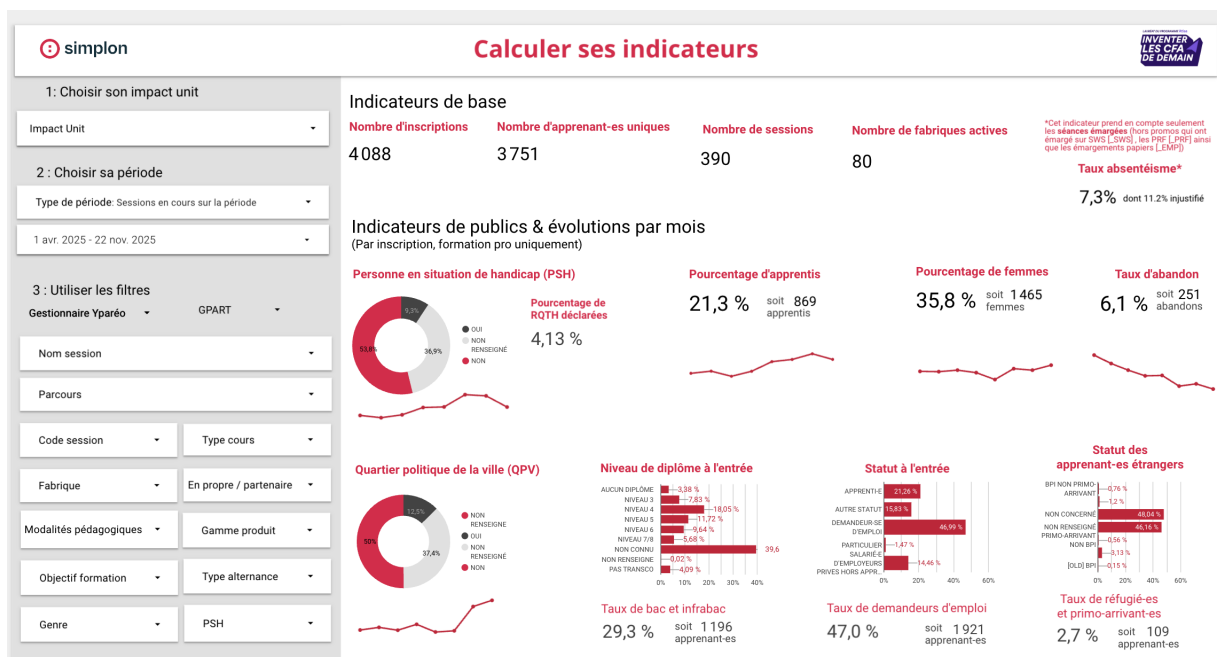
Looker Studio est l'outil principal de visualisation des données chez Simplon.

### 1.2.1 Outils d'analyse et de reporting

Pour répondre aux besoins métiers, plusieurs approches sont utilisées :

#### 1. Tableaux de bord de pilotage :

Ils permettent aux équipes de suivre leurs indicateurs (nombre d'apprenants, sessions, taux de femmes, etc.).





## 2. BigQuery et Excel :

BigQuery héberge les modèles conçus dans DBT. Je l'utilise pour interroger et analyser les données, souvent complétées par des analyses sous Excel (tableaux croisés dynamiques, filtres, regroupements...).

Exemple d'une demande métier :

(a) **Demande – Directrice Régionale Adjointe Simplon Grand Ouest :**

« Bonjour à vous,

Je cherche à faire une extraction des taux de sorties positives de nos formations sur les 18 derniers mois par formation en GDO.

Merci à vous,

Sandrine Charpentier

Directrice Régionale Adjointe - Simplon Grand Ouest.»

(b) **Résolution :**

Après clarification du besoin, j'ai interrogé les modèles BigQuery avec la requête suivante :

```
WITH base AS (
  SELECT * EXCEPT(situation_pro, repondant),
    CASE situation_pro
      WHEN "NON RENSEIGNÉ" THEN NULL
      WHEN "EN ATTENTE" THEN NULL
      WHEN "NON APPLICABLE" THEN NULL
      WHEN "EN RECHERCHE D'EMPLOI" THEN "INACTIF"
      WHEN "SANS ACTIVITÉ" THEN "INACTIF"
      WHEN "INACTIF" THEN "INACTIF"
      WHEN "ACTIF OCCUPÉ HORS ALTERNANCE" THEN "SORTIE POSITIVE"
      WHEN "EN EMPLOI" THEN "SORTIE POSITIVE"
      WHEN "EN FORMATION" THEN "SORTIE POSITIVE"
      ELSE situation_pro
    END AS situation_pro_calcul
  FROM 'fact_sphinx_questionnaires_par_inscription'
)
SELECT DISTINCT
  code_inscription,
  CASE WHEN situation_pro_calcul IS NOT NULL
    THEN 1 ELSE 0 END AS repondant,
  situation_pro_calcul, sortie_positive AS tag_sortie_positive,
  code_session, nom_session, objectif_session, type_produit,
  parcours_formation, date_debut_session, date_fin_session
FROM base
WHERE IUT_session LIKE "%GDO%"
  AND (date_fin_session BETWEEN '2024-05-01' AND '2025-10-31');
```

Les données ont ensuite été analysées dans Google Sheets via un tableau croisé dynamique.

nom_session	Nombre d'apprenants inscrits	Nombre repondans aux questionnaires	Nombre sortie positif	Taux de sortie positive
GDO_ANGERS_CIF_ARTIS_INT_P08	16	8	2	25,00%
GDO_ANGERS_CIF_REFUGIES_142808_SAS	14	5	1	20,00%
GDO_ANGERS_TAI_142816_INT	16	0	0	#DIV/0!
GDO_ANGERS/NANTES_TAI_J0_P5	16	6	4	66,67%
GDO_ANGERS/NANTES_TAI_P07_J0	22	0	0	#DIV/0!
GDO_CAEN_ISEN_DEV IA1_ALT_P1	14	6	3	50,00%
GDO_FABNUMÉRIQUE41_BLOIS_SASCMS_P1_EMP	20	0	0	#DIV/0!
GDO_ISEN_BREST_DEVIA_39232_ALT_P3_EMP	14	1	0	0,00%
GDO_NANTES_CDA_DSI_INT_P11	15	4	3	75,00%
GDO_NANTES_DEV COBOL POEI BNPP_INT_P19	12	4	4	100,00%
GDO_NANTES_POEI REL. CLIENT_INT_P18	13	2	2	100,00%
GDO_NANTES_SAS INFRA_INT_P15	16	9	1	11,11%
GDO_NANTES_SAS_REFUGIES_INT_P17	13	2	1	50,00%
GDO_NANTES_TAI_INT_P14	18	10	8	80,00%
GDO_NANTES_TAI_P20_INT	15	8	7	87,50%
GDO_NANTES_TAI_UP INTERIM_P24_INT	10	0	0	#DIV/0!
GDO_RENNES_DEVFRONT_INT_P15	15	8	3	37,50%
GDO_RENNES_DEVIAP3_ALT_24.P10	13	4	2	50,00%
GDO_RENNES_SAS CMS_FEMMES_INT_P18	17	9	6	66,67%
GDO_RENNES_T.HELPDESK_INT_P16	15	10	6	60,00%
GDO_RENNES_TECHNICIEN HELPDESK_INT_P19	16	0	0	#DIV/0!
GDO_SAINTE-BRIEUC_MSDEV IA_P2_INT_EMP	16	3	2	66,67%
GDO_TOURS GRETA_DEV MSIA3_ALT_P6_EMP	13	0	0	#DIV/0!
GDO_TOURS_GRETA_DEV MSIA3_INT_P6	14	9	9	100,00%
GDO_TOURS_GRETA_DEVIA4_INT_P7_EMP	12	0	0	#DIV/0!
<b>Total général</b>	<b>375</b>	<b>108</b>	<b>64</b>	<b>59,26%</b>

### 1.3 Stratégie de l'équipe DATA

La force de l'équipe DATA réside dans sa rigueur et son organisation, permettant de garantir la fiabilité et la pérennité des informations :

*« Traiter la donnée, c'est plus qu'un travail technique : c'est une responsabilité envers l'information. »*

Trois environnements structurent les projets DATA :

1. **Phase de collecte (RAW) :**

Environnement dédié aux données collectées, données brutes, sans la moindre transformation.

2. **Phase de Développement (DEV) :**

Environnement dédié au développement, aux tests et aux vérifications. Le cycle n'y est pas automatisé.

3. **Phase de Recette (REC) :**

Espace interne à l'équipe DATA permettant de valider les traitements et de disposer d'un secours en cas de problème en production.

4. **Phase de Production (PROD) :**

Environnement final partagé avec l'ensemble de l'entreprise. Les tableaux de bord y sont accessibles aux équipes métiers.

Ce cycle constituait le cadre technique de mon travail quotidien.

**Projet PLACEMENT**

A faire

## Travaux transverses

### 3.1 Résolution des tickets

L'un des rôles les plus importants de l'équipe DATA est de répondre aux différents besoins exprimés par les équipes métiers, notamment les demandes liées aux exports et analyses de données. La plupart du temps, les tableaux de bord existants ne permettent pas aux équipes métiers d'analyser ou d'exporter les données souhaitées. Dans ces situations, c'est à l'équipe DATA de répondre aux demandes dans les meilleurs délais et avec précision.

Accompagné par ma tutrice, j'ai été amené quotidiennement à répondre à ces demandes en investiguant les différents modèles de données ou réadapter les modèles de données existantes ou même construire des nouvelles modèles de données afin de pouvoir répondre à la demande.

Je vais présenter ici en détail quelques demandes que j'ai répondu :

#### 3.1.1 Données certification Agilité

##### Demande

Bonjour,

Dans le cadre du renouvellement de la certification Agilité, j'aurais besoin des données suivantes (entre le 01/06/2021 et le 30/09/2025) : typologie des apprenants certifiés en agilité :

1. Répartition entre hommes et femmes ;
2. Âge moyen.

Merci beaucoup !

Je reste à votre disposition si besoin.

Cheffe de projet qualité et certification

##### Étapes de résolution

#### 1. Compréhension de la demande :

Dans la majorité des cas, les demandes métiers nécessitent une clarification afin d'obtenir un besoin précis et, si nécessaire, de proposer une approche plus adaptée côté DATA. Pour cette demande, les éléments étaient clairs, il n'a donc pas été nécessaire d'organiser un échange supplémentaire.

#### 2. Identification des modèles :

Nous hébergeons un grand nombre de modèles dans *Google Cloud Storage (GCS)*, tous reliés entre eux via des clés uniques. La première difficulté consiste donc à identifier les modèles pertinents pour répondre au besoin.

Les modèles mobilisés pour cette demande sont :

- (a) Le modèle de faits **fact\_certifications\_par\_inscription\_table**, qui fournit les événements de certification pour chaque apprenant.
- (b) Le modèle de dimension **dim\_ypareo\_simplonien**, qui apporte les informations complémentaires apprenant (date de naissance, genre, etc.) nécessaires pour enrichir la table de faits.

### 3. Résolution :

- (a) Extraction des données via requête BigQuery :

```
SELECT DISTINCT
  c.code_simplonien, s.dateNaissance_simplonien,
  DATE_DIFF(c.date_acquisition_competence,
            s.dateNaissance_simplonien, YEAR) AS age,
  c.genre_simplonien, c.nom_session, c.nom_certification,
  c.resultat_certification, c.date_acquisition_competence
FROM 'fact_certifications_par_inscription_table' c
INNER JOIN 'dim_ypareo_simplonien' s
  ON c.code_simplonien = s.code_simplonien
WHERE c.code_certification LIKE "%RS-5487%"
  AND c.resultat_certification = "ACQUIS TOTAL"
  AND c.date_acquisition_competence
    BETWEEN "2021-06-01" AND "2025-09-30"
  AND c.validation_formation NOT LIKE "%ABANDON%"
  AND c.IUT_session NOT LIKE "%TEST999%"
  AND c.resultat_certification NOT LIKE "%NON APPLICABLE%"
  AND c.derniere_inscription = 1
```

- (b) Analyse de la répartition via Excel :

A	B	C	D
		Pour l'année 2025 ( jusqu'à septembre 2025)	
genre_simplonien	Moyenne d'age	Répartition entre Homme et Femme	Répartition (%) entre Homme et Femme
FEMME	34,43023256	86	36,60%
HOMME	36,29530201	149	63,40%
<b>Total général</b>	<b>35,61276596</b>	<b>235</b>	<b>100,00%</b>
		Pour l'année 2024	
genre_simplonien	Moyenne d'age	Répartition entre Homme et Femme	Répartition (%) entre Homme et Femme
FEMME	34,46341463	82	25,95%
HOMME	36,0982906	234	74,05%
<b>Total général</b>	<b>35,67405063</b>	<b>316</b>	<b>100,00%</b>
		Pour l'année 2023	
genre_simplonien	Moyenne d'age	Répartition entre Homme et Femme	Répartition (%) entre Homme et Femme
FEMME	35,60645161	155	24,11%
HOMME	35,00614754	488	75,89%
<b>Total général</b>	<b>35,15085537</b>	<b>643</b>	<b>100,00%</b>
		Pour l'année 2022	
genre_simplonien	Moyenne d'age	Répartition entre Homme et Femme	Répartition (%) entre Homme et Femme
FEMME	33,57894737	39	32,50%
HOMME	29,60759494	81	67,50%
<b>Total général</b>	<b>30,8974359</b>	<b>120</b>	<b>100,00%</b>
		Pour l'année 2021 ( à partir de juin 2021)	
genre_simplonien	Moyenne d'age	Répartition entre Homme et Femme	Répartition (%) entre Homme et Femme
FEMME	34	2	16,67%
HOMME	33,9	10	83,33%
<b>Total général</b>	<b>33,91666667</b>	<b>12</b>	<b>100,00%</b>

### 3.1.2 Apple Reporting

#### Demande

Hello,

Dans le cadre de notre partenariat avec Apple, nous avons reçu une demande concernant le suivi des parcours de nos apprenants. Voici les besoins :

1. Nombre d'apprenants Apple Foundation Program ayant poursuivi une formation Simplon (et si possible lesquelles)
2. Nombre d'apprenants Extended Apple Foundation Program ayant poursuivi une formation Simplon (et si possible lesquelles)

Merci beaucoup !

Je reste à votre disposition si besoin.

Coordinatrice nationale Apple Foundation Program

#### Étapes de résolution

##### 1. Compréhension de la demande :

La demande initiale n'était pas suffisamment précise d'un point de vue DATA. J'ai organisé un échange avec la demandeuse afin de clarifier les objectifs et proposer une structuration adaptée. À l'issue de ce point, nous avons convenu de fournir :

- (a) Le nombre d'apprenants ayant réalisé un AFPE et poursuivi une autre formation.
- (b) Le nombre d'apprenants ayant réalisé un AFP puis poursuivi un AFPE (sans formation intermédiaire).
- (c) Le nombre d'apprenants ayant réalisé un AFP puis poursuivi une autre formation que l'AFPE (en identifiant la première formation suivie après l'AFP).

##### 2. Identification des modèles :

Le modèle mobilisé est : `fact_activite_inscriptions_par_mois_table`, qui décrit mois par mois les activités d'inscription de chaque apprenant.

##### 3. Résolution :

- (a) Extraction des données via requêtes BigQuery.
- (b) Analyse des résultats dans Excel : Analyse de l'export de données en tableau croisé dynamique sur Excel :

A		B	C
		Nombre apprenants ayant fait l'AFPE et ont poursuivi une autre formation ( on a renseigné la première formation suivie après l'AFPE )	
parcours_initial	parcours_fait_apres_AFPE	NB_apprenants	
DÉVELOPPEMENT MOBILE IOS - APPLE FOUNDATION PROGRAM EXTENDED	ADMINISTRATEUR-ICE CLOUD - ECOLE CLOUD MICROSOFT BY SIMPLON	1	
DÉVELOPPEMENT MOBILE IOS - APPLE FOUNDATION PROGRAM EXTENDED	ARTIS	1	
DÉVELOPPEMENT MOBILE IOS - APPLE FOUNDATION PROGRAM EXTENDED	CONCEPTEUR-RICE DÉVELOPPEUR-EUSE D'APPLICATIONS IOS ET OSX - ADVANCED APPLE FOUNDATION PROGRAM	19	
DÉVELOPPEMENT MOBILE IOS - APPLE FOUNDATION PROGRAM EXTENDED	CONCEPTEUR-RICE DÉVELOPPEUR-SE EN ENVIRONNEMENT DEVOPS	2	
DÉVELOPPEMENT MOBILE IOS - APPLE FOUNDATION PROGRAM EXTENDED	DÉVELOPPEMENT MOBILE IOS - APPLE FOUNDATION PROGRAM EXTENDED	1	
DÉVELOPPEMENT MOBILE IOS - APPLE FOUNDATION PROGRAM EXTENDED	DÉVELOPPEUR-EUSE INTÉGRATEUR RICE EN RÉALISATION D'APPLICATIONS WEB	1	
DÉVELOPPEMENT MOBILE IOS - APPLE FOUNDATION PROGRAM EXTENDED	DÉVELOPPEUR-EUSE RÉALITÉ VIRTUELLE/RÉALITÉ AUGMENTÉE - 3D TEMPS RÉEL	1	
DÉVELOPPEMENT MOBILE IOS - APPLE FOUNDATION PROGRAM EXTENDED	DÉVELOPPEUR-SE INTELLIGENCE ARTIFICIELLE	1	
DÉVELOPPEMENT MOBILE IOS - APPLE FOUNDATION PROGRAM EXTENDED	DÉVELOPPEUR-SE INTELLIGENCE ARTIFICIELLE - ECOLE IA MICROSOFT BY SIMPLON	4	
DÉVELOPPEMENT MOBILE IOS - APPLE FOUNDATION PROGRAM EXTENDED	DÉVELOPPEUR-SE WEB ET WEB MOBILE	5	
DÉVELOPPEMENT MOBILE IOS - APPLE FOUNDATION PROGRAM EXTENDED	EXPLOITER LA BLOCKCHAIN DANS LE DÉVELOPPEMENT D'APPLICATIONS	1	
DÉVELOPPEMENT MOBILE IOS - APPLE FOUNDATION PROGRAM EXTENDED	HACKEUSES	2	
DÉVELOPPEMENT MOBILE IOS - APPLE FOUNDATION PROGRAM EXTENDED	MOBILISER LES COMPÉTENCES INFORMATIQUES FONDAMENTALES	5	
DÉVELOPPEMENT MOBILE IOS - APPLE FOUNDATION PROGRAM EXTENDED	SAS DE PRÉQUALIFICATION - FILIÈRE DÉVELOPPEMENT D'APPLICATIONS	1	
DÉVELOPPEMENT MOBILE IOS - APPLE FOUNDATION PROGRAM EXTENDED	SAS DE PRÉQUALIFICATION - FILIÈRE IA MICROSOFT BY SIMPLON	1	
DÉVELOPPEMENT MOBILE IOS - APPLE FOUNDATION PROGRAM EXTENDED	UPSILLING NON-CERTIFIANT	1	
		TOTAL 47	
		Nombre apprenants ayant fait l'AFP et ont poursuivi une AFPE ( sans faire une autre formation intermédiaire )	
parcours_initial	parcours_fait_apres_AFP	NB_apprenants	
DÉVELOPPEMENT MOBILE IOS - APPLE FOUNDATION PROGRAM	DÉVELOPPEMENT MOBILE IOS - APPLE FOUNDATION PROGRAM EXTENDED	203	

Nombre apprenants ayant fait l'AFP et ont poursuivi autre formation que AFPE (on a renseigné la première formation suivie après l'AFP)		
parcours_initial	parcours_fait_apres_AFP	NB_apprenants
R-RICE DÉVELOPPEUR-EUSE D'APPLICATIONS IOS ET OSX - ADVANCED APPLE FOUNDATI	CONCEPTEUR-RICE DÉVELOPPEUR-EUSE D'APPLICATIONS IOS ET OSX	8
DÉVELOPPEMENT MOBILE IOS - APPLE FOUNDATION PROGRAM	ACCESSIBILITÉ NUMÉRIQUE	1
DÉVELOPPEMENT MOBILE IOS - APPLE FOUNDATION PROGRAM	ADMINISTRATEUR-ICE SYSTÈME DEVOPS	1
DÉVELOPPEMENT MOBILE IOS - APPLE FOUNDATION PROGRAM	ADMINISTRATEUR-ICE CLOUD - ECOLE CLOUD MICROSOFT BY SIMPLON	3
DÉVELOPPEMENT MOBILE IOS - APPLE FOUNDATION PROGRAM	ADMINISTRATEUR-ICE D'INFRASTRUCTURES SÉCURISÉES	4
DÉVELOPPEMENT MOBILE IOS - APPLE FOUNDATION PROGRAM	ARTIS	15
DÉVELOPPEMENT MOBILE IOS - APPLE FOUNDATION PROGRAM	ATELIER, SENSIBILISATION, ACCULTURATION	2
DÉVELOPPEMENT MOBILE IOS - APPLE FOUNDATION PROGRAM	AWS RE/START - FONDAMENTAUX DU CLOUD	1
DÉVELOPPEMENT MOBILE IOS - APPLE FOUNDATION PROGRAM	COMPÉTENCES NUMÉRIQUES FONDAMENTALES	7
DÉVELOPPEMENT MOBILE IOS - APPLE FOUNDATION PROGRAM	CONCEPTEUR-RICE DÉVELOPPEUR-EUSE D'APPLICATIONS IOS ET OSX - ADVANCED APPLE FOUNDATION PROGRAM	15
DÉVELOPPEMENT MOBILE IOS - APPLE FOUNDATION PROGRAM	CONCEPTEUR-RICE DÉVELOPPEUR-SE D'APPLICATIONS	18
DÉVELOPPEMENT MOBILE IOS - APPLE FOUNDATION PROGRAM	CONCEPTEUR-RICE DÉVELOPPEUR-SE EN ENVIRONNEMENT DEVOPS	3
DÉVELOPPEMENT MOBILE IOS - APPLE FOUNDATION PROGRAM	CONSULTANT-E DEVOPS	2
DÉVELOPPEMENT MOBILE IOS - APPLE FOUNDATION PROGRAM	DATA ENGINEER	1
DÉVELOPPEMENT MOBILE IOS - APPLE FOUNDATION PROGRAM	DÉVELOPPEMENT MOBILE IOS - APPLE FOUNDATION PROGRAM	93
DÉVELOPPEMENT MOBILE IOS - APPLE FOUNDATION PROGRAM	DÉVELOPPEUR INFORMATIQUE OPTION JEU VIDÉO	1
DÉVELOPPEMENT MOBILE IOS - APPLE FOUNDATION PROGRAM	DÉVELOPPEUR-EUSE INTÉGRATEUR-RICE EN RÉALISATION D'APPLICATIONS WEB	25
DÉVELOPPEMENT MOBILE IOS - APPLE FOUNDATION PROGRAM	DÉVELOPPEUR-EUSE RÉALITÉ VIRTUELLE/RÉALITÉ AUGMENTÉE - 3D TEMPS RÉEL	3
DÉVELOPPEMENT MOBILE IOS - APPLE FOUNDATION PROGRAM	DÉVELOPPEUR-SE CLOUD - ECOLE CLOUD MICROSOFT BY SIMPLON	1
DÉVELOPPEMENT MOBILE IOS - APPLE FOUNDATION PROGRAM	DÉVELOPPEUR-SE DATA	3
DÉVELOPPEMENT MOBILE IOS - APPLE FOUNDATION PROGRAM	DÉVELOPPEUR-SE DATA ET INTELLIGENCE ARTIFICIELLE	8
DÉVELOPPEMENT MOBILE IOS - APPLE FOUNDATION PROGRAM	DÉVELOPPEUR-SE INTELLIGENCE ARTIFICIELLE	3
DÉVELOPPEMENT MOBILE IOS - APPLE FOUNDATION PROGRAM	DÉVELOPPEUR-SE INTELLIGENCE ARTIFICIELLE - ECOLE IA MICROSOFT BY SIMPLON	12
DÉVELOPPEMENT MOBILE IOS - APPLE FOUNDATION PROGRAM	DÉVELOPPEUR-SE WEB ET WEB MOBILE	126
DÉVELOPPEMENT MOBILE IOS - APPLE FOUNDATION PROGRAM	HACKEUSES	28
DÉVELOPPEMENT MOBILE IOS - APPLE FOUNDATION PROGRAM	MÉDIATEUR-ICE NUMÉRIQUE	1
DÉVELOPPEMENT MOBILE IOS - APPLE FOUNDATION PROGRAM	MÉTHODES AGILES DE GESTION ET AMORÇAGE DE PROJET	6
DÉVELOPPEMENT MOBILE IOS - APPLE FOUNDATION PROGRAM	MOBILISER LES COMPÉTENCES INFORMATIQUES FONDAMENTALES	9
DÉVELOPPEMENT MOBILE IOS - APPLE FOUNDATION PROGRAM	OPÉRATEUR-TRICE DE SOLUTIONS DE SÉCURITÉ CLOUD ET HYBRIDE - ECOLE CYBER MICROSOFT BY SIMPLON	2
DÉVELOPPEMENT MOBILE IOS - APPLE FOUNDATION PROGRAM	POEC DATA ANALYST	3
DÉVELOPPEMENT MOBILE IOS - APPLE FOUNDATION PROGRAM	RÉALISER DES APPLICATIONS WEB AVEC UN CMS	11
DÉVELOPPEMENT MOBILE IOS - APPLE FOUNDATION PROGRAM	RÉFÉRENT-E DIGITAL	1
DÉVELOPPEMENT MOBILE IOS - APPLE FOUNDATION PROGRAM	REFUGEEKS	1
DÉVELOPPEMENT MOBILE IOS - APPLE FOUNDATION PROGRAM	SAS DE PRÉQUALIFICATION - FILIÈRE CYBERSÉCURITÉ MICROSOFT BY SIMPLON	5
DÉVELOPPEMENT MOBILE IOS - APPLE FOUNDATION PROGRAM	SAS DE PRÉQUALIFICATION - FILIÈRE DÉVELOPPEMENT D'APPLICATIONS	19
DÉVELOPPEMENT MOBILE IOS - APPLE FOUNDATION PROGRAM	SAS DE PRÉQUALIFICATION - FILIÈRE IA	3
DÉVELOPPEMENT MOBILE IOS - APPLE FOUNDATION PROGRAM	SAS DE PRÉQUALIFICATION - FILIÈRE IA MICROSOFT BY SIMPLON	5
DÉVELOPPEMENT MOBILE IOS - APPLE FOUNDATION PROGRAM	SAS DE PRÉQUALIFICATION - FILIÈRE INFRASTRUCTURES	5
DÉVELOPPEMENT MOBILE IOS - APPLE FOUNDATION PROGRAM	SURVEILLER UN SYSTÈME D'INFORMATION SUR DES CRITÈRES DE SÉCURITÉ INFORMATIQUE	1
DÉVELOPPEMENT MOBILE IOS - APPLE FOUNDATION PROGRAM	TECHNICIEN-NE D'ASSISTANCE INFORMATIQUE	4
DÉVELOPPEMENT MOBILE IOS - APPLE FOUNDATION PROGRAM	TECHNICIEN-NE SUPÉRIEUR SYSTÈMES ET RÉSEAUX	7
DÉVELOPPEMENT MOBILE IOS - APPLE FOUNDATION PROGRAM	TECHNICIEN-NE SUPÉRIEUR SYSTÈMES ET RÉSEAUX - SPÉCIALISATION ANALYSE CYBER	2
DÉVELOPPEMENT MOBILE IOS - APPLE FOUNDATION PROGRAM	TESTEUR JEU VIDÉO ASSURANCE QUALITÉ Q/A	1
	TOTAL	470

## 3.2 Contrôle des données internationales

L'un des plus grands défis de l'équipe internationale consiste à suivre les statistiques de l'entreprise sur ses fabriques à l'international. Chaque pays, notamment en Afrique (ex. : Sénégal), en Europe (ex. : Suisse), au Moyen-Orient (ex. : Liban) et en Amérique latine (ex. : Colombie), renseigne ses données dans un fichier Google Sheet dédié.

L'ensemble de ces fichiers pays est ensuite centralisé dans un unique fichier appelé *Master Sheet*. Ce dernier est utilisé par l'équipe DATA pour construire un tableau de bord permettant à l'équipe internationale de suivre certains indicateurs stratégiques de l'entreprise.

Cette saisie manuelle comporte cependant des risques : erreurs de saisie, oublis de champs, incohérences entre indicateurs ... ect. Ces problèmes dégradent la qualité des données et faussent les indicateurs de pilotage.

### 3.2.1 Données renseignées dans les fichiers pays

Les fichiers de chaque pays contiennent trois grandes catégories de données :

#### Informations de géolocalisation sur la formation :

Région, Pays et ville.

#### Informations sur la formation :

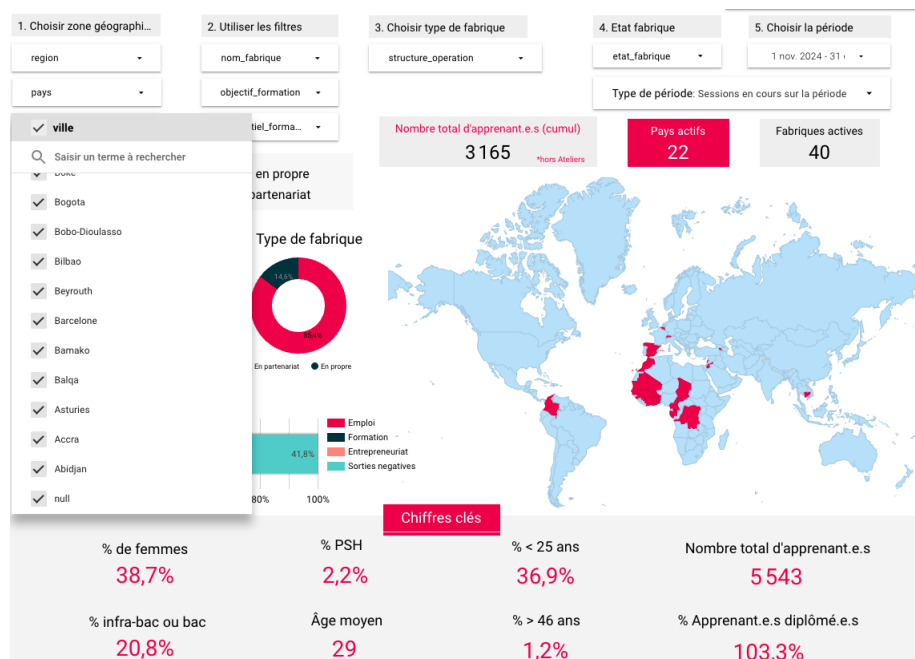
Structure d'opération (Simplon ou partenaire), État de la fabrique (si le centre dde formation est fermée ou active), Public cible (particuliers ou professionnels), Type de formation (métiers, découverte, atelier), Référentiel de formation, Nom de la promotion, Date de début de la formation et Date de fin de la formation.

### Champs d'indicateurs agrégés :

Nombre total d'apprenants au début (hommes + femmes), Dont nombre total de femmes au début, Nombre total d'apprenants ayant abandonné, Nombre total d'apprenants ayant mené la formation à terme, Dont nombre de femmes ayant mené la formation à terme, Nombre d'apprenants diplômés, Nombre de femmes diplômées, Nombre d'apprenants partiellement diplômés, Nombre de femmes partiellement diplômées, ... etc.

#### 3.2.2 Problèmes observés dans le tableau de bord existant

À mon arrivée dans l'équipe DATA, le tableau de bord a été construit directement à partir du fichier *Master Sheet* sans contrôle de la donnée renseignées. Cette méthode a entraîné plusieurs anomalies visibles dans les indicateurs :



- Exemple : un taux de **103,3% de diplômés**, ce qui est un biais. L'erreur provenait des pays ayant renseigné plus de diplômés que d'apprenants ayant mené la formation à terme.
- Certaines valeurs de filtres, comme la ville « Nulle », indiquent que le champ n'a pas été renseigné dans certains fichiers pays.

Ces anomalies faussent fortement les analyses et biaisent la décision.

#### 3.2.3 Initiative de l'équipe Data

On a décidé de mettre en place un système de contrôle automatique permettant :

- d'identifier les incohérences susceptibles de biaiser les indicateurs ;
- d'exclure les lignes incorrectes des analyses ;
- et de remonter les lignes anormales aux équipes internationales afin qu'elles puissent être corrigées rapidement.

#### Technique de résolution

- Côté Data Engineer :** Un flux Talend a été développé pour récupérer quotidiennement de façon automatique l'ensemble des fichiers pays et de les déposer dans **Google Cloud Platform**.
- Mon expertise :**
  - J'ai modélisé les fichiers dans **dbt**, connecté à BigQuery, afin de centraliser l'ensemble des données dans une base unique et contrôlée.



- (b) J'ai implémenté des règles de contrôle de qualité des données et ajouté une colonne permettant d'indiquer automatiquement chaque anomalie détectée sur la ligne correspondante. Les règles mises en place portent notamment sur :

- i. **Fichier pays vide** : aucun enregistrement n'est présent pour le pays concerné. Il s'agit du cas prioritaire : dès qu'il est détecté, aucun autre contrôle n'est réalisé.
- ii. **Format invalide** : un champ numérique ou une date contient une valeur non conforme.
- iii. **Champ obligatoire manquant** : notamment les champs de géolocalisation (région, pays, ville) et les champs nominatifs liés à la formation.
- iv. **Valeur invalide dans un champ nominatif** : une valeur hors liste déroulante est renseignée (ex. type de formation, public cible, état de la fabrique, etc.).
- v. **Incohérence sur les effectifs** : le nombre final ou le nombre d'abandons dépasse le nombre initial.

$$\text{nombre\_final} > \text{nombre\_initial} \quad \text{ou} \quad \text{nombre\_abandons} > \text{nombre\_initial}$$

- vi. **Anomalie sur les validations** : au moins une des conditions suivantes est vraie :

$$\text{nombre\_validations} > \text{nombre\_final}$$

$$\text{nombre\_validation\_partielle} > \text{nombre\_final}$$

$$\text{nombre\_validations} + \text{nombre\_validation\_partielle} > \text{nombre\_final}$$

Cela indique un nombre de validations supérieur au nombre réel d'apprenants présents en fin de parcours.

- vii. **Effectifs féminins incohérents** : les effectifs féminins dépassent les totaux déclarés.

$$\text{nombre\_initial\_femmes} > \text{nombre\_initial}$$

$$\text{nombre\_final\_femmes} > \text{nombre\_final}$$

$$\text{nombre\_validations\_femmes} > \text{nombre\_validations}$$

$$\text{nombre\_validation\_partielle\_femmes} > \text{nombre\_validation\_partielle}$$

$$\text{nombre\_validations\_femmes} > \text{nombre\_final\_femmes}$$

$$\text{nombre\_validation\_partielle\_femmes} > \text{nombre\_final\_femmes}$$

## Résultats obtenus

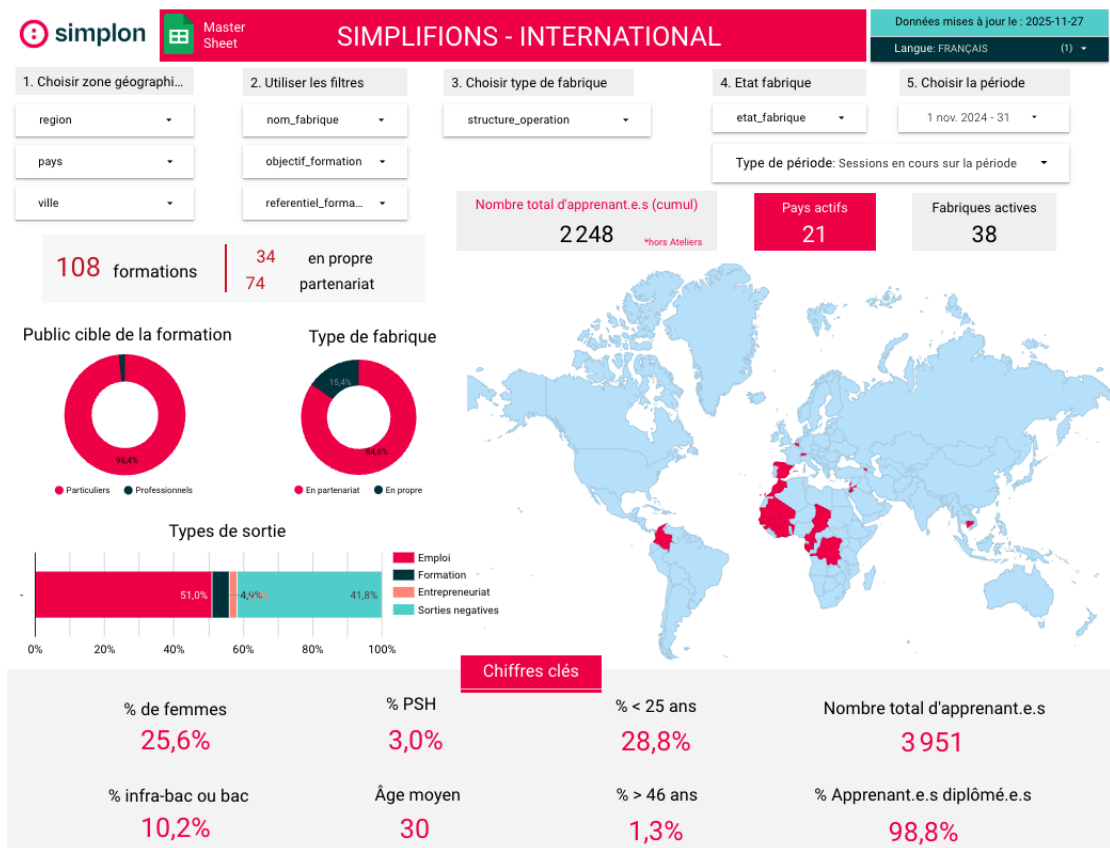
Après la modélisation et la mise en place des contrôles, plusieurs améliorations majeures ont été apportées :

### 1. Adaptation du tableau de bord à la nouvelle modélisation :

Le tableau de bord utilise désormais uniquement les lignes ne présentant aucune anomalie. J'ai également ajouté un bandeau affichant la fraîcheur des données pour chaque pays (collecte quotidienne via Talend, assurée par le Data Engineer). Le bandeau apparaît en **vert** si les données sont récoltées à jour, sinon en **rouge** si elles ne le sont pas, avec l'indication de la dernière mise à jour (ce qui permet aux utilisateurs du tableau de bords de nous alerter en cas de problème de mis à jour de la donnée).

### 2. Fiabilisation des indicateurs :

Les indicateurs sont désormais calculés uniquement à partir des lignes conformes, ce qui garantit des indicateurs fiables et cohérents.



### 3. Création d'une table des anomalies :

J'ai mis en place une table listant toutes les lignes contenant au moins une anomalie ainsi que la nature de l'erreur détectée. Cette table permet aux responsables de saisie d'identifier rapidement les lignes incorrectes et de corriger la donnée dans les fichiers pays :

Message d'anomalie détectée sur la ligne	Région	Pays	Ville	Nom promotion
Au moins un des champs region, ville, pays, nom fabrique, structure operation, etat fabrique, public cible, objectif formation, referentiel formation ou nom de la promotion n'est pas renseigné.	Afrique	Togo		Formateurs Edumérík
Au moins un des champs region, ville, pays, nom fabrique, structure operation, etat fabrique, public cible, objectif formation, referentiel formation ou nom de la promotion n'est pas renseigné.	Afrique	Togo		Edumérík
Au moins un des champs region, ville, pays, nom fabrique, structure operation, etat fabrique, public cible, objectif formation, referentiel formation ou nom de la promotion n'est pas renseigné. et une ou des valeurs de nombre de femmes dépassent les totaux correspondants	Maghreb	Maroc	Tout le Maroc	Projet Kids
une des valeurs de nombre de validations dépasse le nombre final	Afrique	Gabon	Libreville	Ecole 241 Promo 7 Référent Digital
une des valeurs de nombre de validations dépasse le nombre final	Afrique	Gabon	Port-Gentil	Ecole 241 Promo 9 Référent Digital
une des valeurs de nombre de validations dépasse le nombre final	Afrique	Gabon	Libreville	Ecole 241 Promo 10 Développeur Web / Web Mobile
une des valeurs de nombre de validations dépasse le nombre final	Afrique	Gabon	Libreville	Ecole 241 Promo 7 Développeur Web
Au moins un des champs region, ville, pays, nom fabrique, structure operation, etat fabrique, public cible, objectif formation, referentiel formation ou nom de la promotion n'est pas renseigné.	Afrique	Côte d'Ivoire	Abidjan	PSH
Au moins un des champs region, ville, pays, nom fabrique, structure operation, etat fabrique, public cible, objectif formation, referentiel formation ou nom de la promotion n'est pas renseigné.	Afrique	Côte d'Ivoire	Abidjan	P6 DEV
Au moins un des champs region, ville, pays, nom fabrique, structure operation, etat fabrique, public cible, objectif formation, referentiel formation ou nom de la promotion n'est pas renseigné.	Afrique	Côte d'Ivoire	Abidjan	MTN Academy
une ou des valeurs de nombre de femmes dépassent les totaux correspondants	Afrique	Côte d'Ivoire	Abidjan	P1 MONTAGE VIDÉO ET GESTION DE PROJET DIGITAL
Au moins un des champs region, ville, pays, nom fabrique, structure operation, etat fabrique, public cible, objectif formation, referentiel formation ou nom de la promotion n'est pas renseigné.	Afrique	Côte d'Ivoire	Abidjan	PAFESDIA COHORTE 1
Au moins un des champs region, ville, pays, nom fabrique, structure operation, etat fabrique, public cible, objectif formation, referentiel formation ou nom de la promotion n'est pas renseigné. et une des valeurs de nombre de validations dépasse le nombre final	Afrique	Bénin	Cotonou	P1 DCLIC MN Bougelabs Ouidah
Au moins un des champs region, ville, pays, nom fabrique, structure operation, etat fabrique, public cible, objectif formation, referentiel formation ou nom de la promotion n'est pas renseigné. et une des valeurs de nombre de validations dépasse le nombre final	Afrique	Bénin	Cotonou	P1 DCLIC Dev Web Ecole 229 Cotonou
Au moins un des champs region, ville, pays, nom fabrique, structure operation, etat fabrique, public cible, objectif formation, referentiel formation ou nom de la promotion n'est pas renseigné. et une des valeurs de nombre de validations dépasse le nombre final	Afrique	Bénin	Parakou	P1 DCLIC Dev Web Afrikedutech Parakou
une des valeurs de nombre de validations dépasse le nombre final	Afrique	Bénin	Port-Novo	P2 DCLIC MN Simplon

**Conclusion**

La conclusion synthétisant les apports pour l'entreprise et propose des pistes d'évolution ( a faire)