



Data Analyst Stagiaire

---

Rapport d'analyse des données Intercom –  
Dashboard Support

---

Réalisé par :

EL GHARSI Faiza

Promotion : 2024/2025

# Contents

<b>1</b>	<b>Préparation des données et modélisation</b>	<b>1</b>
1.1	Objectif du projet . . . . .	1
1.2	Sources de données . . . . .	1
1.3	Nettoyage et transformation des données . . . . .	2
1.3.1	Prétraitement du fichier <code>CONVERSATIONS.csv</code> et <code>conversation_parts.csv</code> par python . . . . .	2
1.3.2	transformation des données sur Power BI . . . . .	4
1.3.3	Modèle de données (Star Schema) . . . . .	4
<b>2</b>	<b>Création de la base de données et des tables SQL</b>	<b>6</b>
2.1	Table Agents . . . . .	6
2.2	Table <code>CONVERSATIONS_cleaned.csv</code> et <code>CONVERSATIONS_PART_clean.csv</code> .	7
<b>3</b>	<b>Construction du dashboard – Power BI</b>	<b>11</b>
3.1	Chargement des tables . . . . .	11
3.2	Nettoyage dans Power Query . . . . .	12
3.3	Modèle de données (relations) . . . . .	16
3.4	Création des mesures DAX . . . . .	17
3.4.1	Mesure de la satisfaction client (CSAT) . . . . .	17
3.4.2	% de réponses en moins de 5 minutes . . . . .	17
3.4.3	Moments de la semaine où l'équipe est la plus sollicitée . . . . .	18
3.4.4	Détail par personne de l'équipe . . . . .	18
3.5	Construction du dashboard . . . . .	19

# List of Figures

1.1	Parsing du json de <code>CONVERSATIONS.csv</code> . . . . .	3
1.2	Extrait du <code>CONVERSATIONS_cleaned.csv</code> . . . . .	4
1.3	Extrait du <code>conversation_parts_clean.csv</code> . . . . .	4
1.4	Modèle en étoile proposé . . . . .	5
2.1	Création de la table Agents . . . . .	6
2.2	Insertion des premiers agents . . . . .	7
2.3	Importation des fichiers csv pour la création de la table SQL . . . . .	7
2.4	La table <code>CONVERSATIONS_PART_clean.csv</code> . . . . .	8
2.5	La table <code>CONVERSATIONS_cleaned.csv</code> . . . . .	8
2.6	Création de la table <code>CONVERSATIONS_cleaned.csv</code> . . . . .	9
2.7	<code>CONVERSATIONS_PART_clean.csv</code> . . . . .	10
3.1	Connexion de la base de données à power BI . . . . .	11
3.2	Chargement des tables . . . . .	12
3.3	Conversion de type de <code>CREATED_AT</code> . . . . .	13
3.4	Extrait de la table <code>first_user_msg</code> . . . . .	14
3.5	Extrait de colonnes <code>Jour_Semaine</code> et <code>Heure_Journée</code> . . . . .	15
3.6	Modèle de données (relations) . . . . .	16
3.7	Page 1 du dashboard . . . . .	19
3.8	Page 2 du dashboard . . . . .	20

# Listings

3.1	Mesure DAX du score CSAT . . . . .	17
3.2	Colonne calculée du temps de réponse . . . . .	17
3.3	Mesure DAX pour le % de réponses j 5 min . . . . .	17
3.4	Mesure DAX pour le nombre de messages clients . . . . .	18
3.5	Nombre de conversations assignées par agent . . . . .	18
3.6	CSAT moyen par agent . . . . .	18
3.7	Temps de réponse moyen par agent . . . . .	18

# Chapter 1

## Préparation des données et modélisation

### 1.1 Objectif du projet

Lorette, responsable de l'équipe Support, souhaite disposer d'un outil de reporting hebdomadaire basé sur les données issues d'Intercom, permettant de :

- Suivre la satisfaction client (CSAT)
- Évaluer la réactivité de son équipe (temps de première réponse)
- Identifier les périodes de forte sollicitation
- Avoir une vue détaillée par membre de son équipe

### 1.2 Sources de données

Les deux principales sources sont :

- `conversation_parts.csv` : contient l'historique des actions/messages dans chaque conversation.
- `conversations.csv` : contient les métadonnées générales sur les conversations.

Champs clés exploités :

- `Conversation ID` : identifiant commun aux deux tables.
- `Author type` : pour distinguer les messages clients/admin.
- `Created at` : date de création d'un message ou d'une conversation.
- `Rating` : score de satisfaction client (CSAT).
- `Assigned_to / Assignee` : personne à qui la conversation est assignée.

## 1.3 Nettoyage et transformation des données

### 1.3.1 Prétraitement du fichier `CONVERSATIONS.csv` et `conversation_parts.csv` par python

En utilisant google colab, on a développé le script python suivant 1.1 qui effectue une normalisation des données de conversations en extrayant les informations structurées stockées en format JSON dans plusieurs colonnes du fichier original.

- Fonction de sécurité : Création d'une fonction `safe_json_loads()` qui gère les cas problématiques (valeurs nulles, erreurs de format JSON) pour éviter les interruptions du traitement.
- Extraction des évaluations : Décomposition de la colonne `CONVERSATION_RATING` en quatre attributs distincts :
  - Note (`rating`)
  - Commentaire (`remark`)
  - ID du collaborateur (`teammate_id`)
  - Type du collaborateur (`teammate_type`)Traitement des assignations : Extraction depuis la colonne `ASSIGNEE` des informations d'identification (`assignee_id` et `assignee_type`).
- Gestion des tags : Transformation de la colonne `TAGS` contenant des listes d'objets JSON en :
  - Une liste Python de noms de tags (`tags_list`)
  - Une chaîne de caractères concaténée pour faciliter la recherche et le filtrage (`tags_concat`)
- Optimisation du dataset : Suppression des colonnes JSON originales devenues redondantes après extraction.

```

# 1. Charger le CSV
df = pd.read_csv('/content/CONVERSATIONS.csv')

# 2. Parser la colonne CONVERSATION_RATING
conversation_rating = df['CONVERSATION_RATING'].apply(safe_json_loads)

df['rating'] = conversation_rating.apply(lambda x: x.get('rating'))
df['remark'] = conversation_rating.apply(lambda x: x.get('remark'))
df['teammate_id'] = conversation_rating.apply(lambda x: x.get('teammate', {}).get('id'))
df['teammate_type'] = conversation_rating.apply(lambda x: x.get('teammate', {}).get('type'))

# 3. Parser la colonne ASSIGNEE
assignee = df['ASSIGNEE'].apply(safe_json_loads)
df['assignee_id'] = assignee.apply(lambda x: x.get('id'))
df['assignee_type'] = assignee.apply(lambda x: x.get('type'))

# 4. Parser la colonne TAGS (qui est une liste)
tags = df['TAGS'].apply(safe_json_loads)

def extract_tag_names(tag_list):
    if isinstance(tag_list, list):
        return [tag.get('name') for tag in tag_list if 'name' in tag]
    return []

df['tags_list'] = tags.apply(extract_tag_names)
df['tags_concat'] = df['tags_list'].apply(lambda tags: ", ".join(tags) if tags else None)

# 5. Drop les anciennes colonnes JSON si besoin
df_cleaned = df.drop(columns=['CONVERSATION_RATING', 'ASSIGNEE', 'TAGS'])

# 6. Sauvegarder dans un nouveau CSV
df_cleaned.to_csv('conversations_cleaned.csv', index=False)

print(" Nouveau fichier conversations_cleaned.csv généré.")

```

Figure 1.1: Parsing du json de CONVERSATIONS.csv

Ce prétraitement transforme des données semi-structurées complexes en un format tabulaire normalisé, facilitant considérablement les analyses statistiques et la visualisation des données de conversations. On a fait pareil pour `conversation_parts.csv`, vous trouvez le code python complet du parsing sur Git.

Ci-dessous un extrait des deux fichiers `CONVERSATIONS_cleaned.csv` et `conversation_parts_clean.csv` générés après l'exécution du script:

CREATED_AT	ID	OPEN	PRIORITY	READ	STATE	TYPE	UPDATED_AT	WAITING SIN	SDC_BATCHI	SDC_EXTRAC	SDC_RECEIV	SDC_SEQUEI	SDC_TABLE	SNOOZED_UF
2022-01-25 13:42:18.000 Z	53815801339513.0	False	not_priority	True	closed	conversation	2022-02-18 17:54:15.000 Z	2022-02-19 0	2022-02-19 0	2022-02-19 0	1,6452E+18	0		
2022-01-27 13:12:24.000 Z	53815801340062.0	False	not_priority	True	closed	conversation	2022-02-18 17:54:22.000 Z	2022-02-19 0	2022-02-19 0	2022-02-19 0	1,6452E+18	0		
2021-11-22 12:13:52.000 Z	53815801271487.0	False	not_priority	True	closed	conversation	2022-02-18 17:53:49.000 Z	2022-02-19 0	2022-02-19 0	2022-02-19 0	1,6452E+18	0		
2022-01-06 16:40:50.000 Z	53815801327026.0	False	not_priority	True	closed	conversation	2022-02-18 17:54:11.000 Z	2022-02-19 0	2022-02-19 0	2022-02-19 0	1,6452E+18	0		
2021-11-09 08:37:01.000 Z	53815801249855.0	False	not_priority	True	closed	conversation	2022-02-18 17:53:26.000 Z	2022-02-19 0	2022-02-19 0	2022-02-19 0	1,6452E+18	0		
2022-01-25 13:42:18.000 Z	53815801339513.0	False	not_priority	True	closed	conversation	2022-02-18 17:54:15.000 Z	2022-02-19 0	2022-02-19 0	2022-02-19 0	1,6452E+18	0		
2021-11-09 16:39:34.000 Z	53815801054858.0	False	not_priority	True	closed	conversation	2022-02-18 17:53:17.000 Z	2022-02-19 0	2022-02-19 0	2022-02-19 0	1,6452E+18	0		
2022-01-27 13:12:24.000 Z	53815801340062.0	False	not_priority	True	closed	conversation	2022-02-18 17:54:22.000 Z	2022-02-19 0	2022-02-19 0	2022-02-19 0	1,6452E+18	0		
2022-01-22 11:10:14.000 Z	53815801337589.0	False	not_priority	False	closed	conversation	2022-02-18 17:54:12.000 Z	2022-02-19 0	2022-02-19 0	2022-02-19 0	1,6452E+18	0		
2022-01-06 16:01:26.000 Z	53815801327012.0	False	not_priority	True	closed	conversation	2022-02-18 17:54:10.000 Z	2022-02-19 0	2022-02-19 0	2022-02-19 0	1,6452E+18	0		
2021-11-09 10:55:08.000 Z	53815801250050.0	False	priority	True	closed	conversation	2022-02-18 17:53:28.000 Z	2022-02-19 0	2022-02-19 0	2022-02-19 0	1,6452E+18	0		
2022-01-06 12:46:53.000 Z	53815801227029.0	False	not_priority	True	closed	conversation	2022-02-18 17:53:49.000 Z	2022-02-19 0	2022-02-19 0	2022-02-19 0	1,6452E+18	0		
2022-01-06 15:33:21.000 Z	53815801266995.0	False	not_priority	True	closed	conversation	2022-02-18 17:54:10.000 Z	2022-02-19 0	2022-02-19 0	2022-02-19 0	1,6452E+18	0		
2021-11-18 11:54:34.000 Z	53815801267059.0	False	not_priority	True	closed	conversation	2022-02-18 17:53:40.000 Z	2022-02-19 0	2022-02-19 0	2022-02-19 0	1,6452E+18	0		
2021-11-22 09:03:56.000 Z	53815801270432.0	False	not_priority	True	closed	conversation	2022-02-18 17:53:43.000 Z	2022-02-19 0	2022-02-19 0	2022-02-19 0	1,6452E+18	0		
2021-10-01 16:48:26.000 Z	53815801105497.0	False	not_priority	True	closed	conversation	2022-02-18 17:52:43.000 Z	2022-02-19 0	2022-02-19 0	2022-02-19 0	1,6452E+18	0		
2021-10-19 09:35:05.000 Z	53815801159935.0	False	not_priority	True	closed	conversation	2022-02-18 17:53:06.000 Z	2022-02-19 0	2022-02-19 0	2022-02-19 0	1,6452E+18	0		
2021-10-19 07:53:06.000 Z	53815801159902.0	False	not_priority	False	closed	conversation	2022-02-18 17:53:02.000 Z	2022-02-19 0	2022-02-19 0	2022-02-19 0	1,6452E+18	0		
2021-11-09 12:03:50.000 Z	53815801226273.0	False	priority	True	closed	conversation	2022-02-18 17:53:19.000 Z	2022-02-19 0	2022-02-19 0	2022-02-19 0	1,6452E+18	0		
2021-11-29 10:05:47.000 Z	53815801281659.0	False	not_priority	True	closed	conversation	2022-02-18 17:53:53.000 Z	2022-02-19 0	2022-02-19 0	2022-02-19 0	1,6452E+18	0		
2021-11-22 12:13:52.000 Z	53815801271487.0	False	not_priority	True	closed	conversation	2022-02-18 17:53:49.000 Z	2022-02-19 0	2022-02-19 0	2022-02-19 0	1,6452E+18	0		

Figure 1.2: Extrait du CONVERSATIONS\_cleaned.csv

CONVERSATION_CREATED_AT	CONVERSATION_ID	CONVERSATION_CREATED_AT	ID	NOTIFIED_AT	PART_GROUP	TYPE	UPDATED_AT	SDC_BATCHI	SDC_EXTRAC	SDC_RECEIV	SDC_SEQUEI	SDC_TABLE	SNOOZED_UF
2022-01-31 14:02:02.000 Z	5,38158E+13	2022-03-09 11:2022-01-31 1		13007180139	2022-01-31 1: Message	conversation_part	2022-01-31 1:2022-03-10 0:2022-03-10 0:2022-03-						
2022-01-31 14:02:02.000 Z	5,38158E+13	2022-03-09 11:2022-01-31 1		13007122771	2022-01-31 1: Message	conversation_part	2022-01-31 1:2022-03-10 0:2022-03-10 0:2022-03-						
2022-01-31 14:02:02.000 Z	5,38158E+13	2022-03-09 11:2022-01-31 1		13009945647	2022-01-31 1: Message	conversation_part	2022-01-31 1:2022-03-10 0:2022-03-10 0:2022-03-						
2022-01-31 14:02:02.000 Z	5,38158E+13	2022-03-09 11:2022-01-31 1		13008753050	2022-01-31 1: Message	conversation_part	2022-01-31 1:2022-03-10 0:2022-03-10 0:2022-03-						
2022-01-31 14:02:02.000 Z	5,38158E+13	2022-03-09 11:2022-01-31 1		13009940447	2022-01-31 1: Close	conversation_part	2022-01-31 1:2022-03-10 0:2022-03-10 0:2022-03-						
2022-01-31 14:02:02.000 Z	5,38158E+13	2022-03-09 11:2022-01-31 1		13007074663	2022-01-31 1: Quick Reply	conversation_part	2022-01-31 1:2022-03-10 0:2022-03-10 0:2022-03-						
2022-01-31 14:02:02.000 Z	5,38158E+13	2022-03-09 11:2022-01-31 1		13007400685	2022-01-31 1: Message	conversation_part	2022-01-31 1:2022-03-10 0:2022-03-10 0:2022-03-						
2022-01-31 14:02:02.000 Z	5,38158E+13	2022-03-09 11:2022-01-31 1		13007128281	2022-01-31 1: Message	conversation_part	2022-01-31 1:2022-03-10 0:2022-03-10 0:2022-03-						
2022-01-31 14:02:02.000 Z	5,38158E+13	2022-03-09 11:2022-01-31 1		13009799278	2022-01-31 1:16:02:22.000 Z	conversation_part	2022-01-31 1:2022-03-10 0:2022-03-10 0:2022-03-						
2022-01-31 14:02:02.000 Z	5,38158E+13	2022-03-09 11:2022-01-31 1		13007073038	2022-01-31 1: Assignment	conversation_part	2022-01-31 1:2022-03-10 0:2022-03-10 0:2022-03-						
2022-01-31 14:02:02.000 Z	5,38158E+13	2022-03-09 11:2022-01-31 1		13009679851	2022-01-31 1: Message	conversation_part	2022-01-31 1:2022-03-10 0:2022-03-10 0:2022-03-						
2022-01-31 14:02:02.000 Z	5,38158E+13	2022-03-09 11:2022-03-09 1		13633563435	2022-03-09 10:10:16.000 Z	conversation_part	2022-03-09 11:2022-03-10 0:2022-03-10 0:2022-03-						
2022-01-31 14:02:02.000 Z	5,38158E+13	2022-03-09 11:2022-03-09 1		13633564039	2022-03-09 11: Close	conversation_part	2022-03-09 11:2022-03-10 0:2022-03-10 0:2022-03-						
2022-01-31 14:02:02.000 Z	5,38158E+13	2022-03-09 11:2022-01-31 1		13007108912	2022-01-31 1: Message	conversation_part	2022-01-31 1:2022-03-10 0:2022-03-10 0:2022-03-						
2022-01-31 14:02:02.000 Z	5,38158E+13	2022-03-09 11:2022-01-31 1		13007395704	2022-01-31 1: Message	conversation_part	2022-01-31 1:2022-03-10 0:2022-03-10 0:2022-03-						
2022-01-31 14:02:02.000 Z	5,38158E+13	2022-03-09 11:2022-01-31 1		13008411179	2022-01-31 1: Message	conversation_part	2022-01-31 1:2022-03-10 0:2022-03-10 0:2022-03-						
2022-01-31 14:02:02.000 Z	5,38158E+13	2022-03-09 11:2022-01-31 1		13007074034	2022-01-31 1: Message	conversation_part	2022-01-31 1:2022-03-10 0:2022-03-10 0:2022-03-						
2022-01-31 14:02:02.000 Z	5,38158E+13	2022-03-09 11:2022-01-31 1		13008951519	2022-01-31 1: Message	conversation_part	2022-01-31 1:2022-03-10 0:2022-03-10 0:2022-03-						
2022-01-31 14:02:02.000 Z	5,38158E+13	2022-03-09 11:2022-01-31 1		13007119285	2022-01-31 1: Assignment	conversation_part	2022-01-31 1:2022-03-10 0:2022-03-10 0:2022-03-						
2022-01-31 14:02:02.000 Z	5,38158E+13	2022-03-09 11:2022-01-31 1		13007109545	2022-01-31 1: Message	conversation_part	2022-01-31 1:2022-03-10 0:2022-03-10 0:2022-03-						
2022-01-31 14:02:02.000 Z	5,38158E+13	2022-03-09 11:2022-01-31 1		13007078711	2022-01-31 1: Message	conversation_part	2022-01-31 1:2022-03-10 0:2022-03-10 0:2022-03-						

Figure 1.3: Extrait du conversation\_parts\_clean.csv

## 1.3.2 transformation des données sur Power BI

Les étapes suivantes ont été réalisées dans Power BI / Power Query, on va les détailler plus dans la partie power BI :

Conversion des dates au bon format datetime.

- Filtrage : suppression des doublons et des lignes dont les champs essentiels sont vides.
- Filtrage des messages de type "bot" (non pertinents pour l'analyse).
- Création de colonnes :
  - Jour\_semaine, Heure à partir des dates des messages.
  - Time\_to\_first\_response pour calculer le délai entre le 1er message client et la 1re réponse admin.
- Création de tables dérivées :
  - First\_user\_msg, First\_admin\_msg, jointes par conversation.

## 1.3.3 Modèle de données (Star Schema)

Pour répondre aux exigences de reporting de Lorette, on a adopté une modélisation en étoile, en séparant les dimensions analytiques (Temps, Agents) des faits (Conversations, Messages). Ce modèle favorise la lisibilité, les calculs DAX efficaces, et une analyse rapide

par agent et période. Ainsi, le schéma en étoile suivant 1.4 a été conçu (en utilisant lucidchart) pour optimiser les analyses :

- Table de faits principale (**Faits\_Conversations**): Contient une ligne par conversation, pour les mesures agrégées. Utilisé pour calculer : CSAT moyen, % de réponses < 5 min et Nombre total de conversations.
- Table de faits secondaires (**Faits\_Messages**): Contient tous les messages (**conversation\_parts**). Utilisé pour compter les messages, analyser les heures de sollicitation et distinguer les messages agents / utilisateurs.
- Dimension (**Dim\_Temps**): Utile pour les filtres temporels et les moments de sollicitation.
- Dimension (**Dim\_Agents**): Pour analyser les performances par agent.

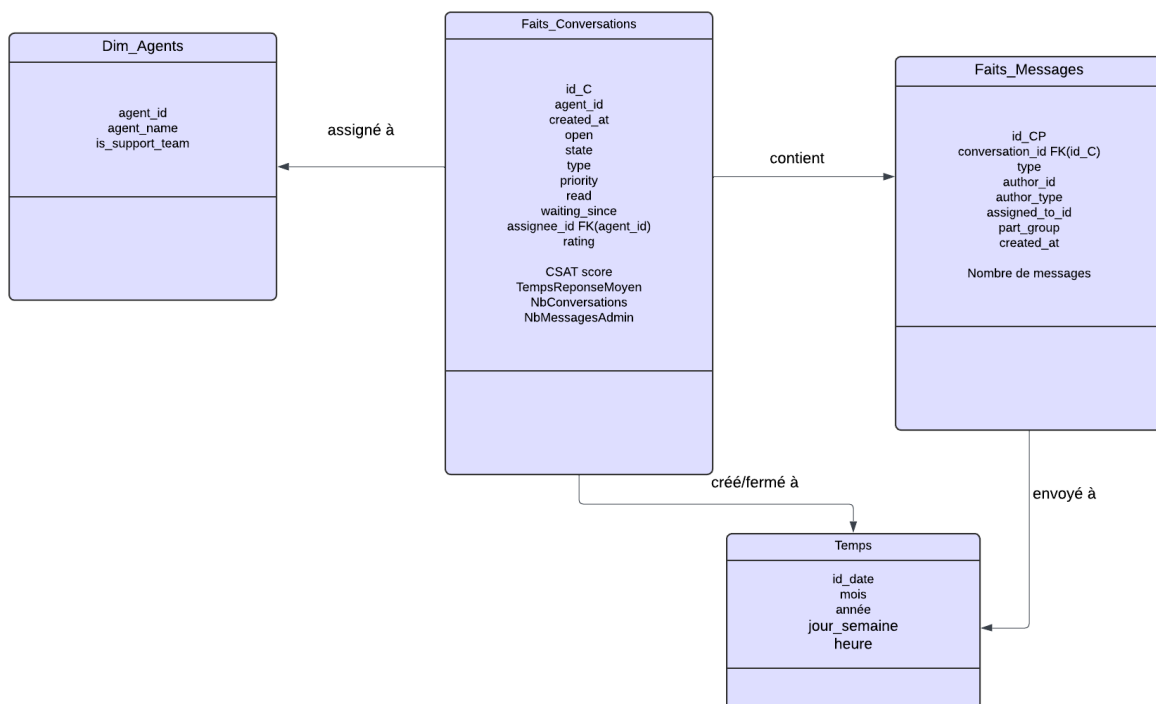


Figure 1.4: Modèle en étoile proposé



## Chapter 2

# Création de la base de données et des tables SQL

En utilisant MySQL, on a créé une base de données skello. Puis, on a créé les tables.

### 2.1 Table Agents

La table Agents regroupe les membres de l'équipe support. Sa création a été réalisée avec l'instruction SQL suivante :

```
CREATE TABLE Agents (  
    agent_id int NOT NULL,  
    agent_name VARCHAR(255),  
    is_support_team BOOLEAN,  
    primary key(agent_id)  
);
```

Figure 2.1: Création de la table Agents

Les premiers agents ont été insérés manuellement dans la base à l'aide de la commande suivante :

```
INSERT INTO Agents (agent_id, agent_name, is_support_team) VALUES
('5217337', 'H  loise', true),
('5391224', 'Justine', true),
('5440474', 'Patrick', true),
('5300290', 'Raphael', true);
```

Figure 2.2: Insertion des premiers agents

## 2.2 Table CONVERSATIONS\_cleaned.csv et CONVERSATIONS\_PART\_clean.csv

Apr  s l'ex  cution du script Python, on obtient deux fichiers : `CONVERSATIONS_cleaned.csv` et `CONVERSATIONS_PART_clean.csv`. Ces fichiers peuvent   tre import  s directement dans MySQL pour la cr  ation de leurs tables SQL    travers Table data import wizard.

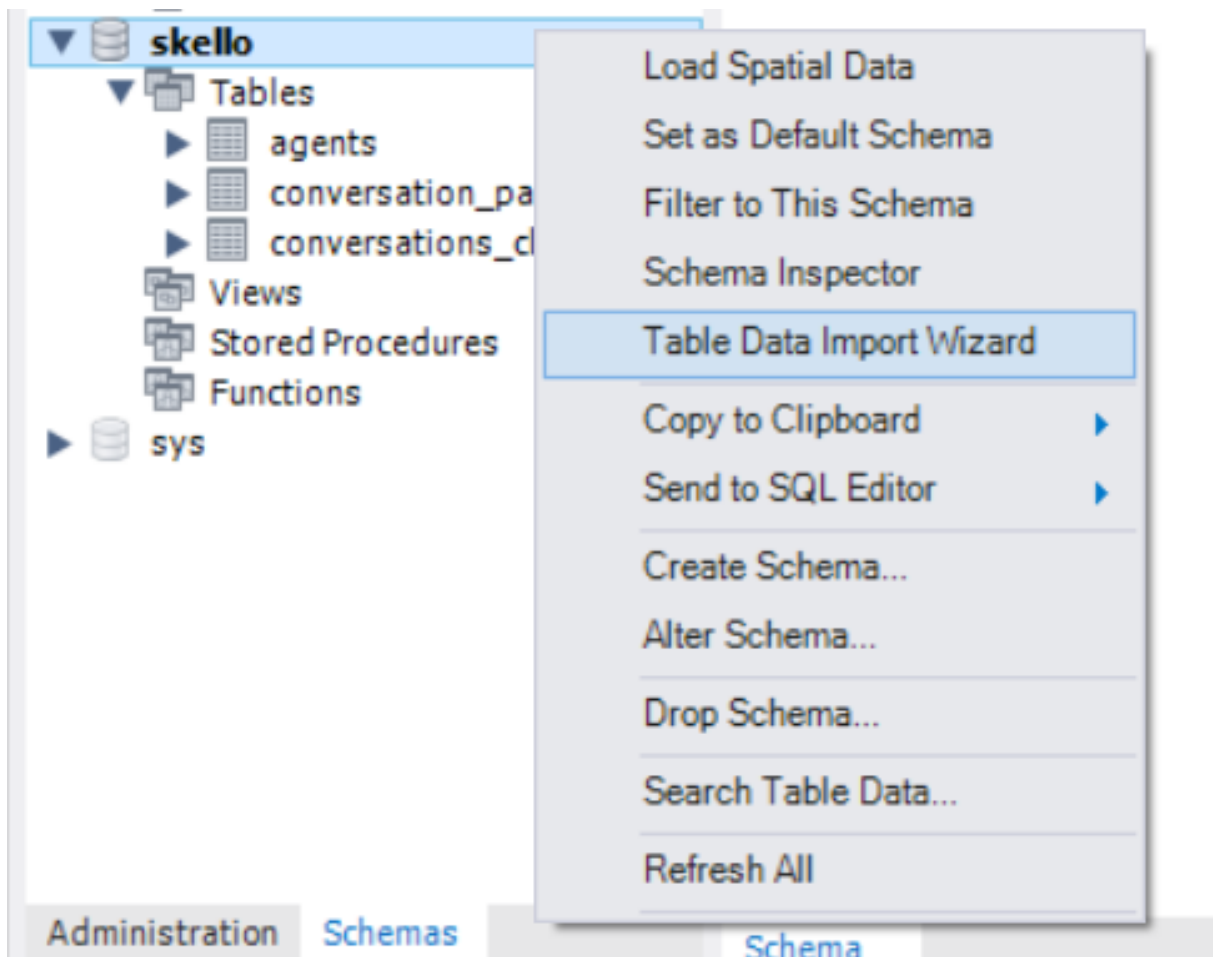


Figure 2.3: Importation des fichiers csv pour la cr  ation de la table SQL

Ci-dessous un extrait des deux tables SQL obtenues:

	CONVERSATION_CREATED_AT	CONVERSATION_ID	CONVERSATION_UPDATED_AT	CREATED_AT	ID	NOTIFIED_AT	PART_GROUP	TYPE	^
▶	2022-01-31 14:02:02.000 Z	53815801349742	2022-03-09 10:10:18.000 Z	2022-01-31 14:06:26.000 Z	13007180139	2022-01-31 14:06:26.000 Z	Message	conversati	
	2022-01-31 14:02:02.000 Z	53815801349742	2022-03-09 10:10:18.000 Z	2022-01-31 14:04:07.000 Z	13007122771	2022-01-31 14:04:07.000 Z	Message	conversati	
	2022-01-31 14:02:02.000 Z	53815801349742	2022-03-09 10:10:18.000 Z	2022-01-31 16:08:44.000 Z	13009945647	2022-01-31 16:08:44.000 Z	Message	conversati	
	2022-01-31 14:02:02.000 Z	53815801349742	2022-03-09 10:10:18.000 Z	2022-01-31 15:15:22.000 Z	13008753050	2022-01-31 15:15:22.000 Z	Message	conversati	
	2022-01-31 14:02:02.000 Z	53815801349742	2022-03-09 10:10:18.000 Z	2022-01-31 16:08:30.000 Z	13009940447	2022-01-31 16:08:30.000 Z	Close	conversati	
	2022-01-31 14:02:02.000 Z	53815801349742	2022-03-09 10:10:18.000 Z	2022-01-31 14:02:08.000 Z	13007074663	2022-01-31 14:02:08.000 Z	Quick Reply	conversati	
	2022-01-31 14:02:02.000 Z	53815801349742	2022-03-09 10:10:18.000 Z	2022-01-31 14:16:05.000 Z	13007400685	2022-01-31 14:16:05.000 Z	Message	conversati	
	2022-01-31 14:02:02.000 Z	53815801349742	2022-03-09 10:10:18.000 Z	2022-01-31 14:04:19.000 Z	13007128281	2022-01-31 14:04:19.000 Z	Message	conversati	
	2022-01-31 14:02:02.000 Z	53815801349742	2022-03-09 10:10:18.000 Z	2022-01-31 16:02:22.000 Z	13009799278	2022-01-31 16:02:22.000 Z		conversati	
	2022-01-31 14:02:02.000 Z	53815801349742	2022-03-09 10:10:18.000 Z	2022-01-31 14:02:04.000 Z	13007073038	2022-01-31 14:02:04.000 Z	Assignment	conversati	
	2022-01-31 14:02:02.000 Z	53815801349742	2022-03-09 10:10:18.000 Z	2022-01-31 15:57:12.000 Z	13009679851	2022-01-31 15:57:12.000 Z	Message	conversati	
	2022-01-31 14:02:02.000 Z	53815801349742	2022-03-09 10:10:18.000 Z	2022-03-09 10:10:16.000 Z	13633563435	2022-03-09 10:10:16.000 Z		conversati	
	2022-01-31 14:02:02.000 Z	53815801349742	2022-03-09 10:10:18.000 Z	2022-03-09 10:10:19.000 Z	13633564039	2022-03-09 10:10:19.000 Z	Close	conversati	
	2022-01-31 14:02:02.000 Z	53815801349742	2022-03-09 10:10:18.000 Z	2022-01-31 14:03:33.000 Z	13007108912	2022-01-31 14:03:33.000 Z	Message	conversati	
	2022-01-31 14:02:02.000 Z	53815801349742	2022-03-09 10:10:18.000 Z	2022-01-31 14:15:53.000 Z	13007395704	2022-01-31 14:15:53.000 Z	Message	conversati	
	2022-01-31 14:02:02.000 Z	53815801349742	2022-03-09 10:10:18.000 Z	2022-01-31 15:00:42.000 Z	13008411179	2022-01-31 15:00:42.000 Z	Message	conversati	
	2022-01-31 14:02:02.000 Z	53815801349742	2022-03-09 10:10:18.000 Z	2022-01-31 14:02:07.000 Z	13007074034	2022-01-31 14:02:07.000 Z	Message	conversati	
	2022-01-31 14:02:02.000 Z	53815801349742	2022-03-09 10:10:18.000 Z	2022-01-31 15:24:30.000 Z	13008951519	2022-01-31 15:24:30.000 Z	Message	conversati	
	2022-01-31 14:02:02.000 Z	53815801349742	2022-03-09 10:10:18.000 Z	2022-01-31 14:03:58.000 Z	13007119285	2022-01-31 14:03:58.000 Z	Assignment	conversati	
	2022-01-31 14:02:02.000 Z	53815801349742	2022-03-09 10:10:18.000 Z	2022-01-31 14:03:35.000 Z	13007109545	2022-01-31 14:03:35.000 Z	Message	conversati	▼

Figure 2.4: La tableCONVERSATIONS\_PART\_clean.csv

	CREATED_AT	ID	OPEN	PRIORITY	READ	STATE	TYPE	UPDATED_AT	WAITING_SINCE	_SDC_BATCHED_AT	_SDC
▶	2022-01-25 13:42:18.000 Z	53815801339513	False	not_priority	True	closed	conversation	2022-02-18 17:54:15.000 Z		2022-02-19 04:48:59.049 Z	2022
	2022-01-25 13:42:18.000 Z	53815801339513	False	not_priority	True	closed	conversation	2022-02-18 17:54:15.000 Z		2022-02-19 04:48:59.049 Z	2022
	2022-01-22 11:10:14.000 Z	53815801337589	False	not_priority	False	closed	conversation	2022-02-18 17:54:12.000 Z		2022-02-19 04:48:59.041 Z	2022
	2022-01-06 16:01:26.000 Z	53815801327012	False	not_priority	True	closed	conversation	2022-02-18 17:54:10.000 Z		2022-02-19 04:48:58.960 Z	2022
	2021-11-22 09:03:56.000 Z	53815801270432	False	not_priority	True	closed	conversation	2022-02-18 17:53:43.000 Z		2022-02-19 04:48:58.946 Z	2022
	2021-10-01 16:48:26.000 Z	53815801105497	False	not_priority	True	closed	conversation	2022-02-18 17:52:43.000 Z		2022-02-19 04:48:58.916 Z	2022
	2021-11-29 10:05:47.000 Z	53815801281659	False	not_priority	True	closed	conversation	2022-02-18 17:53:53.000 Z		2022-02-19 04:48:58.952 Z	2022
	2022-01-06 16:20:01.000 Z	53815801327017	False	not_priority	True	closed	conversation	2022-02-18 17:54:10.000 Z		2022-02-19 04:48:58.960 Z	2022
	2021-11-22 11:55:27.000 Z	53815801271479	False	not_priority	True	closed	conversation	2022-02-18 17:53:51.000 Z		2022-02-19 04:48:58.951 Z	2022
	2022-01-06 12:18:16.000 Z	53815801326921	False	not_priority	True	closed	conversation	2022-02-18 17:54:06.000 Z		2022-02-19 04:48:58.958 Z	2022
	2021-11-18 11:16:26.000 Z	53815801267038	False	not_priority	True	closed	conversation	2022-02-18 17:53:39.000 Z		2022-02-19 04:48:58.944 Z	2022
	2021-11-18 06:00:23.000 Z	53815801265729	False	not_priority	True	closed	conversation	2022-02-18 17:53:35.000 Z		2022-02-19 04:48:58.942 Z	2022
	2021-10-19 10:07:46.000 Z	53815801159948	False	not_priority	True	closed	conversation	2022-02-18 17:53:08.000 Z		2022-02-19 04:48:58.928 Z	2022
	2021-11-17 10:06:58.000 Z	53815801264058	False	not_priority	True	closed	conversation	2022-02-18 17:53:33.000 Z		2022-02-19 04:48:58.941 Z	2022
	2022-01-05 09:50:04.000 Z	53815801323017	False	not_priority	True	closed	conversation	2022-02-18 17:54:04.000 Z		2022-02-19 04:48:58.957 Z	2022
	2022-01-27 14:26:16.000 Z	53815801340092	False	not_priority	True	closed	conversation	2022-02-18 17:54:21.000 Z		2022-02-19 04:48:59.062 Z	2022
	2022-01-27 16:32:01.000 Z	53815801340157	False	not_priority	True	closed	conversation	2022-02-18 17:54:25.000 Z		2022-02-19 04:48:59.067 Z	2022
	2021-11-18 09:56:18.000 Z	53815801266897	False	not_priority	True	closed	conversation	2022-02-18 17:53:38.000 Z		2022-02-19 04:48:58.943 Z	2022
	2021-10-25 11:39:05.000 Z	53815801211600	False	not_priority	True	closed	conversation	2022-02-18 17:53:18.000 Z		2022-02-19 04:48:58.934 Z	2022
	2021-10-19 07:23:13.000 Z	53815801159898	False	not_priority	True	closed	conversation	2022-02-18 17:53:01.000 Z		2022-02-19 04:48:58.925 Z	2022

Figure 2.5: La tableCONVERSATIONS\_cleaned.csv

Sinon si on veut créer les tables en utilisant les requêtes SQL, voici le script utilisé:

```
CREATE TABLE conversations_cleaned (  
    CREATED_AT DATETIME,  
    ID DOUBLE PRIMARY KEY,  
    OPEN VARCHAR(10),  
    PRIORITY VARCHAR(50),  
    READ VARCHAR(10),  
    STATE VARCHAR(50),  
    TYPE VARCHAR(50),  
    UPDATED_AT DATETIME,  
    WAITING_SINCE DATETIME,  
    _SDC_BATCHED_AT DATETIME,  
    _SDC_EXTRACTED_AT DATETIME,  
    _SDC_RECEIVED_AT DATETIME,  
    _SDC_SEQUENCE BIGINT,  
    _SDC_TABLE_VERSION INT,  
    ASSIGNEE TEXT,  
    SNOOZED_UNTIL DATETIME,  
    rating VARCHAR(10),  
    remark TEXT,  
    teammate_type VARCHAR(50),  
    teammate_id DOUBLE,  
    assignee_id INT,
```

---

Figure 2.6: Création de la table CONVERSATIONS\_cleaned.csv

```

> CREATE TABLE conversation_part_clean (
    AUTHOR_id VARCHAR(255),
    AUTHOR_type VARCHAR(100),
    CONVERSATION_CREATED_AT DATETIME,
    CONVERSATION_ID BIGINT,
    CONVERSATION_UPDATED_AT DATETIME,
    CREATED_AT DATETIME,
    ID BIGINT PRIMARY KEY,
    NOTIFIED_AT DATETIME,
    PART_GROUP VARCHAR(100),
    TYPE VARCHAR(50),
    UPDATED_AT DATETIME,
    _SDC_BATCHED_AT DATETIME,
    _SDC_EXTRACTED_AT DATETIME,
    _SDC_RECEIVED_AT DATETIME,
    _SDC_SEQUENCE BIGINT,
    _SDC_TABLE_VERSION INT
);

```

Figure 2.7: CONVERSATIONS\_PART\_clean.csv

# Chapter 3

## Construction du dashboard – Power BI

### 3.1 Chargement des tables

D'abord, on a commencé par connecter notre base de données à power BI.



Figure 3.1: Connexion de la base de données à power BI

Puis, on a chargé les tables qu'on a besoin pour les visualisations demandées.

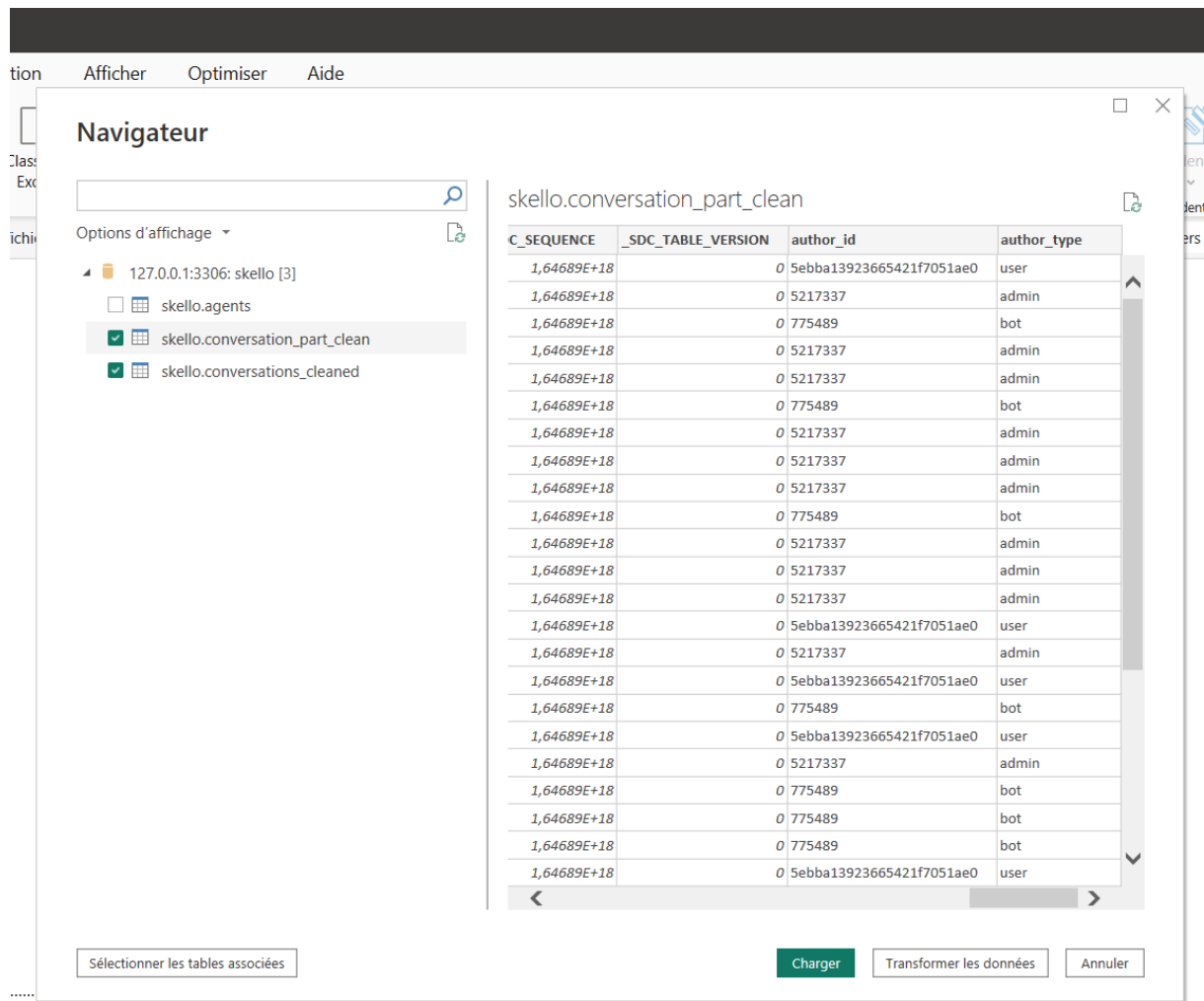


Figure 3.2: Chargement des tables

## 3.2 Nettoyage dans Power Query

Afin d'assurer la qualité et la cohérence des analyses futures, un important travail de nettoyage a été réalisé dans Power Query avant la modélisation. Ce nettoyage a concerné les deux tables principales issues de notre base de données nommée sqello : **CONVERSATIONS\_cleaned** et **CONVERSATIONS\_PART\_clean**.

### 1. Conversion des types de données

De nombreuses colonnes contenant des timestamps (comme **CREATED\_AT**, **UPDATED\_AT**, etc.) étaient initialement au format texte. Elles ont été converties au format **Date/Heure** pour permettre des calculs temporels précis (par exemple : temps de réponse, durée de traitement, etc.).


 CREATED_AT	<sup>12</sup> <sub>3</sub> ID
31/01/2022 15:06:26	13007180139
31/01/2022 15:04:07	13007122771
31/01/2022 16:15:22	13008753050
31/01/2022 15:16:05	13007400685
31/01/2022 15:04:19	13007128281
31/01/2022 16:57:12	13009679851
31/01/2022 15:03:33	13007108912
31/01/2022 15:15:53	13007395704
31/01/2022 16:00:42	13008411179
31/01/2022 16:24:30	13008951519
31/01/2022 15:02:15	13007077347
31/01/2022 15:49:43	13008163021
31/01/2022 16:12:00	13008676846
23/11/2021 11:40:48	11880487080
23/11/2021 11:42:06	11880506708
23/11/2021 11:40:12	11880478012
23/11/2021 11:52:34	11880669548

Figure 3.3: Conversion de type de CREATED\_AT

## 2. Filtrage des données inutiles

Les messages de type *bot* ont été exclus de la table `conversation_parts`, conformément à l'énoncé, afin d'éviter de fausser les calculs de volume ou de temps de réponse.

De plus, seules les lignes de type "message" ont été conservées, car elles concernent des échanges humains (utilisateurs ou admins).

## 3. Normalisation des champs textuels

Par exemple, la colonne `RATING`, initialement stockée sous forme de texte (ex :



"5.0"), a été convertie en **nombre décimal**. Les valeurs vides ou nulles ont été conservées pour ne pas biaiser le calcul du score CSAT.

#### 4. Détection et suppression des doublons

Des doublons ont été détectés dans certaines colonnes censées être uniques (par exemple : ID = 53815801339513 dans `conversations_cleaned`). Ces lignes ont été supprimées ou filtrées manuellement afin de garantir la validité des relations entre les tables du modèle de données.

#### 5. Création de tables intermédiaires

Deux tables ont été créées :

- `first_user_msg` : contenant, pour chaque conversation, le premier message envoyé par l'utilisateur.
- `first_admin_msg` : contenant, pour chaque conversation, la première réponse envoyée par un membre de l'équipe support.

Requêtes [5]

skello conversations\_cleaned  
skello conversation\_part\_clean  
skello agents  
**first\_user\_msg**  
first\_admin\_msg

= Table.TransformColumnTypes("#Lignes filtrées3",{{"time\_to\_first\_response", type number}})

	CONVERSATION_ID	user_first_msg_time	first_admin_msg.admin_first_msg_time	1.2 time_to_first_response
1	5,38158E+13	31/01/2022 15:02:15	31/01/2022 15:04:07	1,866666667
2	5,38158E+13	23/11/2021 11:40:12	23/11/2021 11:40:48	0,6
3	5,38158E+13	08/10/2021 14:00:23	08/10/2021 14:01:31	1,133333333
4	5,38158E+13	09/11/2021 09:12:07	09/11/2021 09:30:07	18
5	5,38158E+13	22/11/2021 13:14:00	22/11/2021 13:17:08	3,133333333
6	5,38158E+13	25/01/2022 14:43:23	25/01/2022 14:55:45	12,366666667
7	5,38158E+13	08/10/2021 15:51:16	08/10/2021 15:52:31	1,25
8	5,38158E+13	01/10/2021 18:52:50	01/10/2021 18:59:11	6,35
9	5,38158E+13	19/10/2021 11:02:52	19/10/2021 11:12:28	9,6
10	5,38158E+13	19/10/2021 10:45:45	19/10/2021 10:54:58	9,216666667
11	5,38158E+13	03/01/2022 17:37:04	03/01/2022 17:40:58	3,9
12	5,38158E+13	24/01/2022 20:23:53	25/01/2022 10:25:27	841,566666667
13	5,38158E+13	25/10/2021 10:43:55	25/10/2021 10:49:11	5,266666667
14	5,38158E+13	22/11/2021 13:34:55	22/11/2021 13:39:03	4,133333333
15	5,38158E+13	18/11/2021 09:52:57	18/11/2021 10:54:23	61,433333333
16	5,38158E+13	27/01/2022 15:40:59	27/01/2022 15:42:28	1,483333333
17	5,38158E+13	18/11/2021 08:49:20	18/11/2021 09:07:12	17,866666667
18	5,38158E+13	18/11/2021 09:09:35	18/11/2021 09:10:44	1,15
19	5,38158E+13	27/01/2022 17:52:12	27/01/2022 17:53:28	1,266666667
20	5,38158E+13	25/10/2021 11:11:04	25/10/2021 11:17:11	6,116666667
21	5,38158E+13	25/10/2021 10:04:27	25/10/2021 10:07:35	3,133333333
22	5,38158E+13	27/01/2022 18:27:08	27/01/2022 18:29:45	2,616666667
23	5,38158E+13	19/10/2021 11:09:59	19/10/2021 11:15:55	5,933333333
24	5,38158E+13	25/01/2022 16:17:07	25/01/2022 16:43:52	26,75
25	5,38158E+13	16/12/2021 11:43:53	16/12/2021 11:55:01	11,133333333

Figure 3.4: Extrait de la table `first_user_msg`

#### 6. Création de colonnes calculées

Plusieurs colonnes personnalisées ont été ajoutées pour faciliter l'analyse, notamment :

- `Jour_Semaine` : extraite depuis `CREATED_AT` pour identifier les jours de forte activité.
- `Heure_Journée` : pour repérer les pics horaires de sollicitation.
- `time_to_first_response` : mesurant le temps entre le premier message client et la première réponse d'un agent.

$A_C^B$ Jour_Semaine	$1_3^2$ Heure_Journée
lundi	15
lundi	15
lundi	16
lundi	15
lundi	15
lundi	16
lundi	15
lundi	15
lundi	16
lundi	16
lundi	15
lundi	15
lundi	16
mardi	11
mardi	11
mardi	11
mardi	11
mardi	11
mardi	11
mardi	11
mardi	11
mardi	11
lundi	12
lundi	9
lundi	9

### 3.3 Modèle de données (relations)

Une fois les données nettoyées dans Power Query, un **modèle relationnel** a été conçu dans Power BI pour structurer l'analyse et faciliter la création de mesures DAX pertinentes.

1. **Table conversation\_parts\_clean**

Cette table constitue le cœur du modèle et contient les différentes actions réalisées dans les conversations, telles que les messages envoyés par les utilisateurs ou les admins.

2. **Table conversations\_cleaned**

Cette table contient les métadonnées des conversations. Elle est reliée à `conversation_parts_clean` via la clé `CONVERSATION_ID`, selon une relation *plusieurs-à-un* (many-to-one).

3. **Table Agents**

Cette table référence les membres de l'équipe support (Héloïse, Justine, Patrick, Raphaël), en précisant leur `agent_id`, leur nom, et un booléen `is_support_team` pour faciliter les filtres. Elle est liée à `conversations_cleaned` via `assignee_id`.

4. **Autres dimensions techniques**

Des tables auxiliaires ont été ajoutées pour des calculs spécifiques :

- `first_user_msg` : identifie le premier message de l'utilisateur par conversation.
- `first_admin_msg` : identifie le premier message d'un admin par conversation.

Ces deux tables permettent de calculer le *temps de réponse initial*, nécessaire au KPI "Réponses en  $\leq$  5 minutes".

5. **Relations dans Power BI**

Toutes les relations sont en cardinalité "*plusieurs-à-un*" avec *filtrage croisé en direction unique*, pour préserver la logique descendante des agrégations.

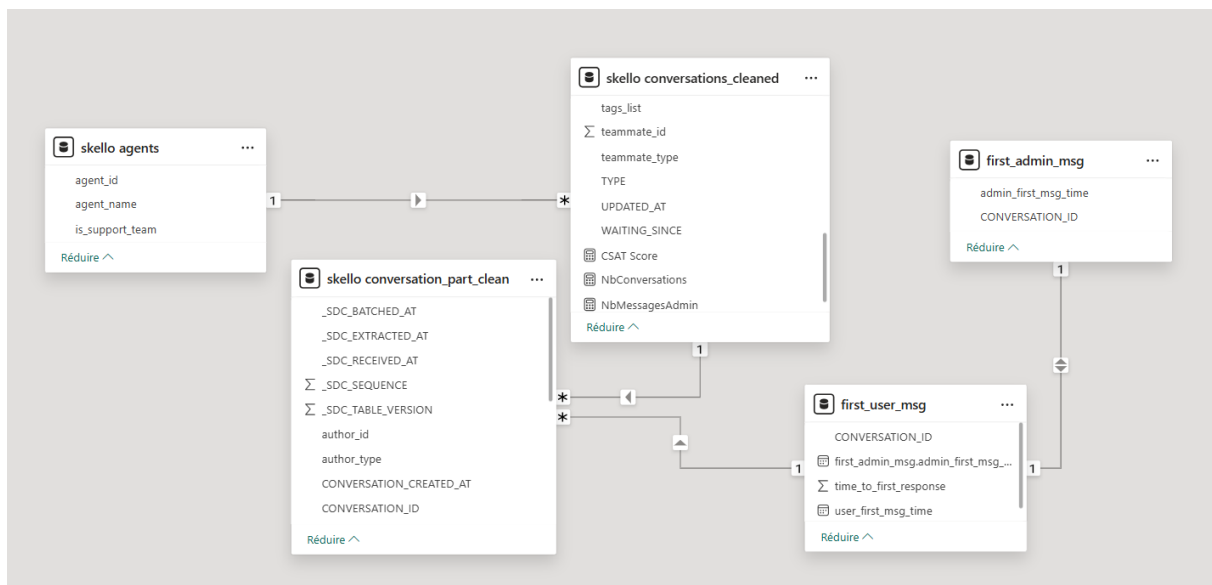


Figure 3.6: Modèle de données (relations)

## 3.4 Création des mesures DAX

Dans le but de suivre les indicateurs clés de performance demandés par Lorette, plusieurs mesures DAX ont été créées dans Power BI. Ces mesures permettent d'extraire de manière dynamique des informations utiles pour l'analyse hebdomadaire de l'activité de l'équipe Support. Voici le détail des mesures développées :

### 3.4.1 Mesure de la satisfaction client (CSAT)

**Objectif :** mesurer le niveau moyen de satisfaction client à travers les notes laissées sur les conversations.

**Prétraitement :** la colonne `rating` de la table `conversations_cleaned` était initialement au format texte. Un nettoyage a été effectué dans Power Query en ajoutant une nouvelle colonne `CSAT_Note` pour convertir cette colonne en nombre décimal, en filtrant les valeurs nulles ou vides.

**Mesure DAX :**

```
1 CSAT Score = AVERAGE('skello conversations_cleaned'[CSAT_Note])
```

Listing 3.1: Mesure DAX du score CSAT

### 3.4.2 % de réponses en moins de 5 minutes

**Objectif :** mesurer la réactivité de l'équipe en identifiant les conversations où le premier message admin a été envoyé moins de 5 minutes après le premier message client.

**Préparation :**

Deux tables auxiliaires ont été créées dans Power Query :

- `first_user_msg` : première date d'un message envoyé par l'utilisateur.
- `first_admin_msg` : première date d'un message envoyé par un admin.

Une table combinée `first_response_table` a été générée pour calculer la différence entre les deux timestamps (temps de réponse initial).

**Colonne calculée :**

```
1 time_to_first_response =  
2 DATEDIFF(  
3     first_response_table[user_first_msg_time],  
4     first_response_table[admin_first_msg_time],  
5     SECOND  
6 )
```

Listing 3.2: Colonne calculée du temps de réponse

**Mesure DAX :**

```
1 TauxR_ponseMoins5min =  
2 VAR TotalConv = COUNTROWS('first_user_msg')  
3 VAR ConvRapides =  
4     CALCULATE(  
5         COUNTROWS('first_user_msg'),
```

```

6         'first_user_msg'[time_to_first_response] < 5
7     )
8 RETURN
9 DIVIDE(ConvRapides, TotalConv, 0)

```

Listing 3.3: Mesure DAX pour le % de réponses j 5 min

### 3.4.3 Moments de la semaine où l'équipe est la plus sollicitée

**Objectif :** identifier les jours et heures avec le plus grand volume de messages clients afin d'optimiser la planification des ressources.

**Colonnes ajoutées dans Power Query :**

- **Jour de la semaine :** à partir de la colonne `CREATED_AT` de `conversation_parts_clean`, via :

Transformer > Date > Jour de la semaine > Nom du jour

- **Heure du jour :** extraite également à partir de `CREATED_AT`, via :

Transformer > Heure > Heure

**Mesure DAX :**

```

1 Nombre de messages = CALCULATE(COUNTROWS('skello conversation_part_clean
2     '),
3     FILTER('skello conversation_part_clean',
4         'skello conversation_part_clean'[PART_GROUP] = "message" &&
5         'skello conversation_part_clean'[author_type] = "user"))

```

Listing 3.4: Mesure DAX pour le nombre de messages clients

### 3.4.4 Détail par personne de l'équipe

**Objectif :** comparer les performances de chaque agent du Support en termes de volume traité, satisfaction client (CSAT) et temps de réponse.

**Mesures développées :**

- **Nombre de conversations assignées :**

```

1 NbConversations = COUNT('skello conversations_cleaned'[ID])
2

```

Listing 3.5: Nombre de conversations assignées par agent

- **CSAT moyen par agent :**

```

1 CSAT Score = AVERAGE('skello conversations_cleaned'[CSAT_Note])
2

```

Listing 3.6: CSAT moyen par agent

- **Temps de réponse moyen (en min) :**

```

1 TempsReponseMoyen = AVERAGE('first_user_msg'[time_to_first_response
2     ])

```

Listing 3.7: Temps de réponse moyen par agent

### 3.5 Construction du dashboard

Une fois les données nettoyées, les relations établies entre les tables, et les mesures DAX créées, la dernière étape a été de concevoir un dashboard lisible, interactif et pertinent, répondant aux besoins de Lorette pour son meeting hebdomadaire de pilotage de l'équipe Support.

L'objectif principal était de synthétiser les KPIs (indicateurs clés) relatifs à la qualité, la rapidité de réponse et la répartition de charge de l'équipe.

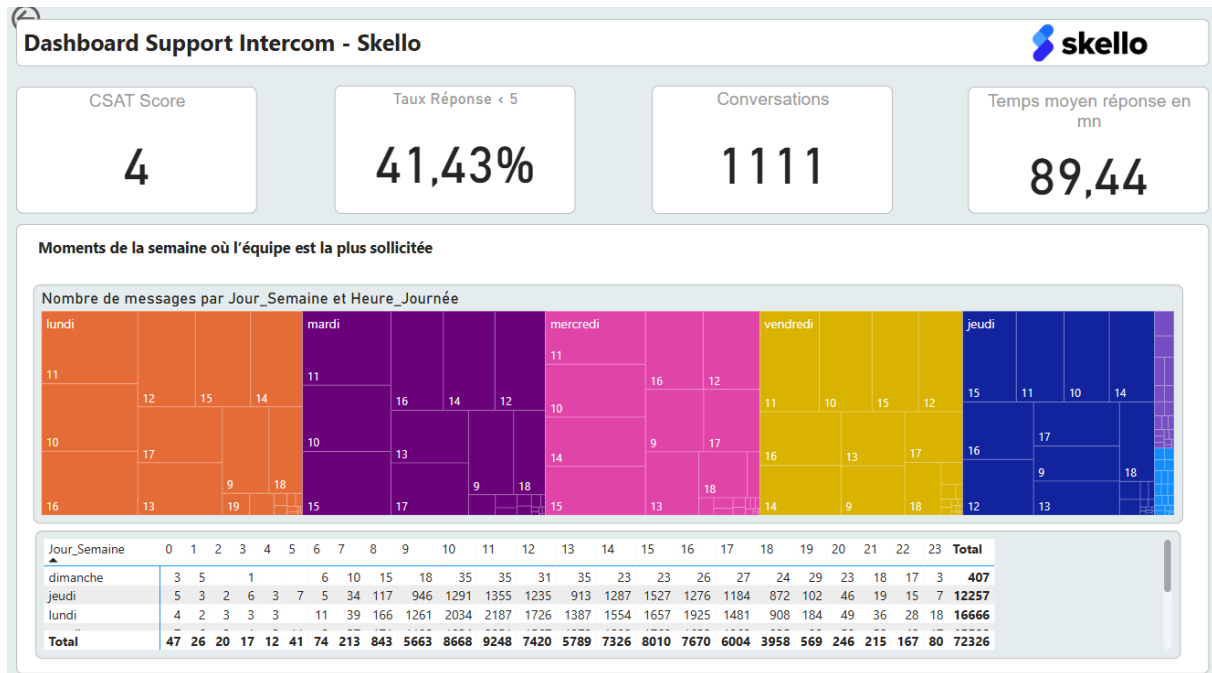
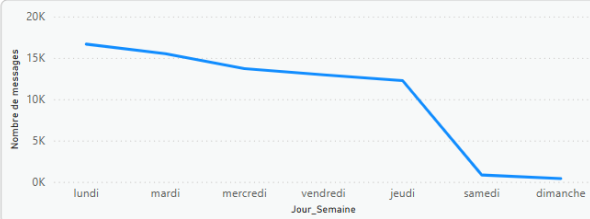


Figure 3.7: Page 1 du dashboard

## Performance de l'équipe

agent_name	NbConversations	NbMessagesAdmin	CSAT Score	TempsReponseMoyen
	765	51571	4	89,44
Héloïse	341	1461	4	89,44
Raphael	5	16	4	89,44
Justine				89,44
Patrick				89,44
<b>Total</b>	<b>1111</b>	<b>53048</b>	<b>4</b>	<b>89,44</b>

## Nombre de messages par Jour\_Semaine



## Objectifs de la semaine

- Améliorer le taux de réponse < 5min à 45% (actuellement 41,43%)
- Maintenir la CSAT au-dessus de 4/5
- Réduire le temps de réponse moyen à <85min
- Former Patrick et Justine pour renforcer l'équipe

Figure 3.8: Page 2 du dashboard