



# INSEA

Préparé par : ***M. DARIF Yassine***

*Année universitaire : 2024–2025*



## Résumé

Ce projet a pour objectif principal d'analyser les avis clients des agences bancaires au Maroc, extraits de la plateforme Google Maps. En tirant parti d'une architecture de données moderne (Modern Data Stack), nous avons mis en œuvre un pipeline complet allant de la collecte et du nettoyage des données, à leur transformation et enfin à leur visualisation. L'analyse des sentiments et l'identification des thèmes dominants ont permis de dégager des tendances significatives concernant la satisfaction client et les points faibles ou forts des agences, fournissant ainsi des informations actionnables pour les institutions bancaires.

## Abstract

This project aims to analyze customer reviews of Moroccan bank branches collected from Google Maps. By leveraging a modern data stack, we developed a comprehensive pipeline encompassing data collection, cleaning, transformation, and visualization. Sentiment analysis and topic modeling techniques were applied to extract valuable insights into customer satisfaction and identify key strengths or weaknesses of the branches. The generated actionable insights can serve as a valuable resource for banking institutions seeking to improve their services.

## Table des matières

Résumé .....	3
Abstract.....	3
Liste des Figures.....	5
1 Introduction Générale .....	6
2 Présentation du Projet .....	8
2.1 Objectifs .....	8
2.2 Problématique .....	8
2.3 Architecture du Système .....	9
3 Phase 1 : Collecte des Données .....	10
3.1 Source de données.....	10
3.2 Champs Collectés par le Scraping .....	11
3.3 Automatisation.....	12
3.4 Défis rencontrés et solutions.....	12
4 Phase 2 : Nettoyage et Transformation des Données.....	13
4.1 Nettoyage des Données .....	13
4.2 Enrichissement des Données .....	13
5 Phase 3 : Modélisation des Données.....	15
5.1 Conception du Data Mart .....	15
5.2 Chargement et Orchestration .....	16
6 Phase 4 : Analyse des Données et Reporting .....	17
6.1 Connexion des Données via Aiven .....	17
6.2 Construction de Tableaux de Bord Interactifs .....	17
7 Stack Technologique Détaillée .....	22
8 Perspectives d'Amélioration .....	23
9 Conclusion.....	24

## Liste des Figures

Figure 1: Architecture du système de données.....	9
Figure 2 : Image de Google Maps .....	10
Figure 3: Champs Collectés par le Scraping.....	12
Figure 4 : Airflow Graph .....	12
Figure 5: Données Brutes .....	14
Figure 6 : Données transformées.....	14
Figure 7 : Schéma ER du modèle en étoile.....	15
Figure 8 : Modèle DBT de transformation .....	16
Figure 9 : Connexion des Données via Aiven.....	17
Figure 10 : Tendance des sentiments par banque et par agence.....	18
Figure 11 : Sujets les plus positifs et les plus négatifs .....	19
Figure 12 : Classement de performance des agences .....	20
Figure 13 : Indicateurs d'expérience client.....	21

# 1 Introduction Générale

Dans un monde où la digitalisation est omniprésente, la voix du client est devenue une ressource inestimable, façonnant la réputation et le succès des entreprises. Les plateformes en ligne, telles que Google Maps, sont désormais des dépositaires d'une mine d'informations non structurées sous forme d'avis, reflétant les expériences réelles des consommateurs. Pour le secteur bancaire, traditionnellement perçu comme conservateur, l'écoute et l'analyse de ces retours clients sont devenues cruciales pour maintenir un avantage concurrentiel, améliorer les services et renforcer la fidélité. Ce projet s'inscrit précisément dans cette dynamique, en cherchant à transformer ces vastes volumes de données textuelles en informations exploitables.

Le défi majeur réside dans la capacité à extraire, traiter et interpréter ces avis de manière efficace. Comment les institutions bancaires peuvent-elles naviguer dans cette masse de commentaires pour identifier les tendances émergentes, comprendre le sentiment général des clients et détecter les problèmes récurrents affectant leurs agences ou leurs services ? La réponse à cette problématique réside dans le développement d'une solution d'ingénierie des données robuste et intelligente, capable de collecter automatiquement ces avis, de les nettoyer, de les analyser sémantiquement, puis de les présenter sous une forme compréhensible et actionable.

L'objectif principal de ce projet est de concevoir et de mettre en œuvre un pipeline de données complète qui automatise le cycle de vie des avis clients de Google Maps. Cela inclut la collecte des données via des techniques de web scraping avancées, leur nettoyage et transformation pour les rendre analysables, l'application de méthodes de traitement du langage naturel (NLP) pour l'analyse de sentiment et l'extraction thématique, la modélisation et le stockage des données dans une base de données optimisée, et enfin la visualisation interactive des insights via des tableaux de bord dynamiques. Notre méthodologie s'appuiera sur une approche itérative, tirant parti des meilleures pratiques en matière d'ingénierie des données et d'analyse de texte, en utilisant des technologies de pointe telles que Python, Apache Airflow, DBT, PostgreSQL et Looker Studio.

Ce rapport est structuré pour vous guider à travers toutes les phases de ce projet. Après cette introduction générale, nous présenterons le cas d'usage spécifique et l'intérêt de cette initiative pour le secteur bancaire. Nous explorerons ensuite les travaux connexes qui ont influencé notre approche, avant de détailler le cahier des charges. Les chapitres suivants seront consacrés à

l'architecture générale du projet, aux technologies utilisées et à une description détaillée de chaque phase du pipeline : de la collecte des données au nettoyage, à la modélisation et à l'analyse/visualisation. Nous aborderons également les défis rencontrés et les solutions apportées, présenterons les résultats clés et les interprétations, et enfin, discuterons des perspectives d'amélioration et de la conclusion générale du projet.

## 2 Présentation du Projet

Le projet "Analyzing Customer Reviews of Bank Agencies in Morocco using a Modern Data Stack" est une initiative visant à extraire, traiter et analyser les avis clients des agences bancaires marocaines disponibles sur Google Maps.

### 2.1 Objectifs

- **Centraliser les avis clients** : Créer une base de données unifiée et structurée des avis collectés, facilitant ainsi leur exploitation.
- **Analyser les sentiments** : Déterminer la polarité (positive, négative, neutre) des avis pour évaluer le niveau général de satisfaction client.
- **Identifier les thèmes dominants** : Détecter les sujets les plus fréquemment abordés par les clients (e.g., temps d'attente, accueil, services en ligne, guichets automatiques) afin de cerner les préoccupations majeures.
- **Évaluer la performance des agences** : Classer les agences en fonction de leurs notes moyennes et des sentiments exprimés, permettant ainsi d'identifier les agences performantes et celles nécessitant une attention particulière.
- **Fournir des insights actionnables** : Générer des rapports et des tableaux de bord interactifs pour les décideurs, leur permettant de prendre des décisions éclairées basées sur les retours clients.

### 2.2 Problématique

Comment pouvons-nous transformer les milliers d'avis clients non structurés, éparpillés sur Google Maps, en indicateurs de performance clairs et en informations stratégiques exploitables pour les banques opérant au Maroc ? Le défi majeur est de passer de données brutes et hétérogènes à des insights actionnables, permettant d'évaluer la satisfaction client, d'identifier les points faibles des services, et d'orienter les décisions opérationnelles et marketing de manière scalable et moderne. Il s'agit de comprendre comment un pipeline de données avancée peut révéler des tendances cachées et des problèmes récurrents, offrant ainsi aux banques les outils nécessaires pour améliorer continuellement leur offre et leur réputation en ligne au Maroc.



## 2.3 Architecture du Système

Le système mis en place pour ce projet repose sur une architecture de données moderne, communément appelée "Modern Data Stack". Cette architecture est conçue pour la scalabilité, la flexibilité et la facilité d'intégration. Elle se décompose en plusieurs étapes clés :

- **Extraction** : La collecte des données est effectuée via des techniques de scraping web (BeautifulSoup, Selenium) pour récupérer les avis depuis Google Maps.
- **Stockage** : Les données brutes et transformées sont stockées dans une base de données relationnelle robuste (PostgreSQL).
- **Transformation** : Les données brutes sont nettoyées, enrichies et modélisées à l'aide de DBT (Data Build Tool), qui permet de définir des transformations SQL modulaires et versionnées.
- **Orchestration** : Le pipeline de données, de la collecte à la transformation, est automatisé et planifié via Apache Airflow.
- **Analyse et Visualisation** : Les données modélisées sont exploitées dans un outil de Business Intelligence (Looker Studio) pour créer des tableaux de bord interactifs et des rapports.

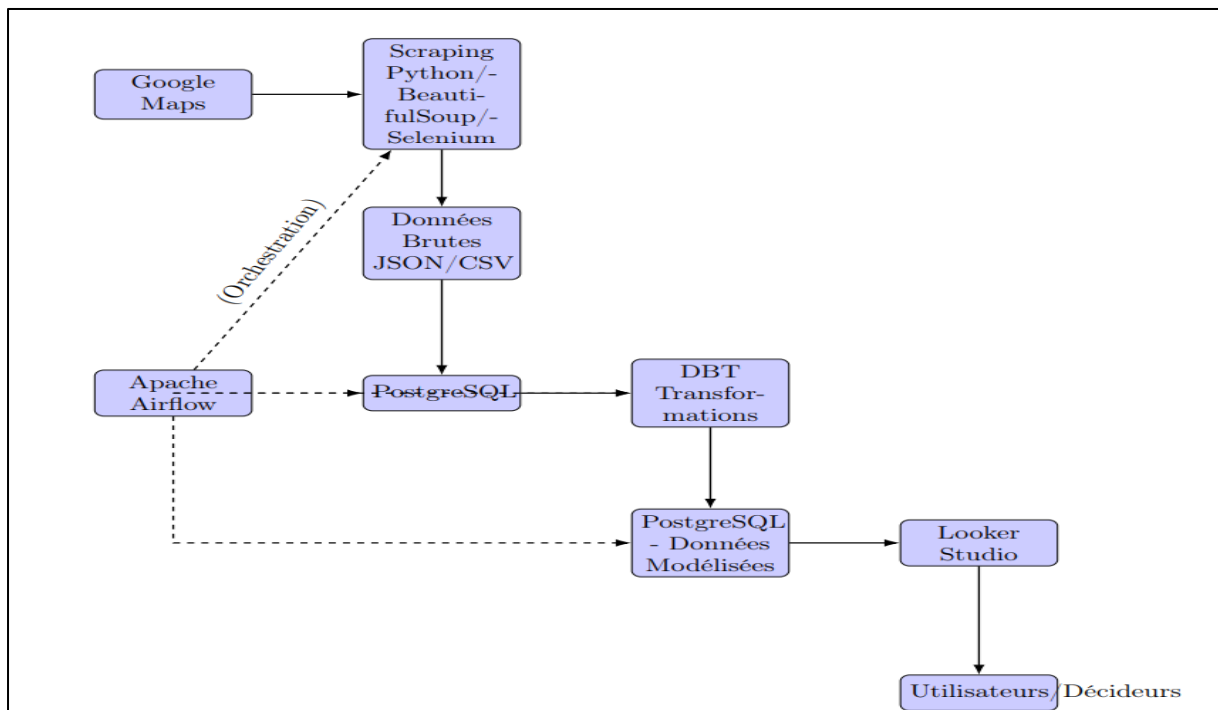


Figure 1: Architecture du système de données

### 3 Phase 1 : Collecte des Données

Cette phase cruciale vise à extraire les avis clients depuis Google Maps.

#### 3.1 Source de données

Google Maps, une plateforme riche en avis générés par les utilisateurs

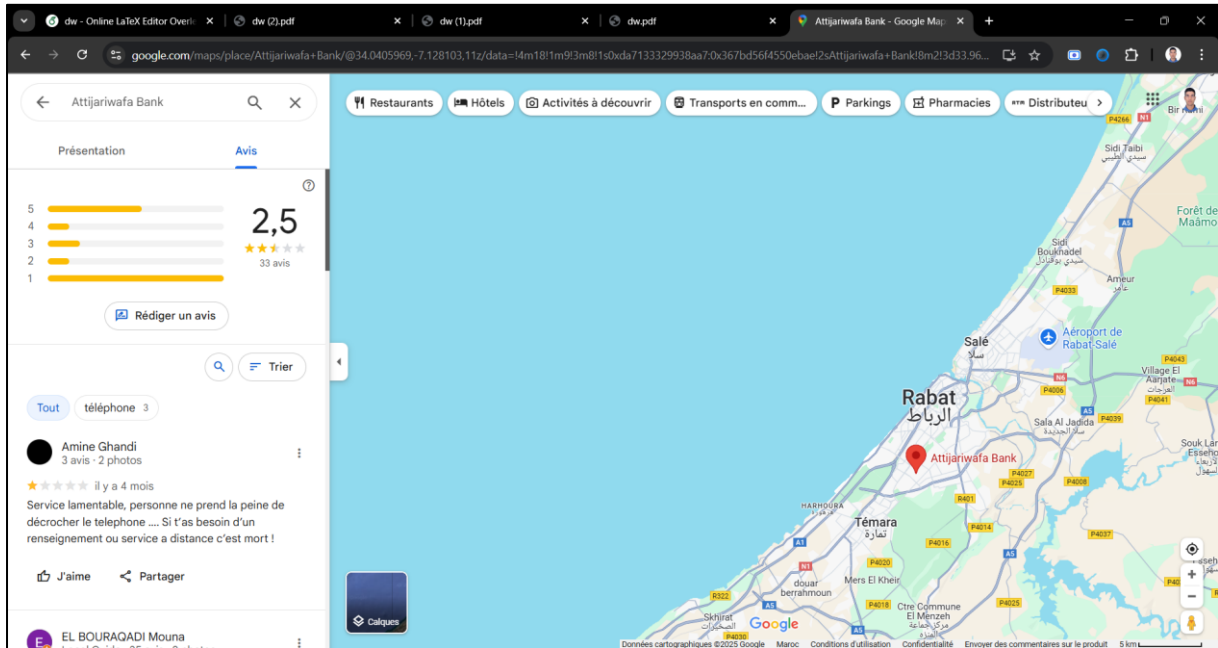


Figure 2 : Image de Google Maps

#### Outils et Méthodologie

- **Python** : Langage de programmation principal pour le développement des scripts de scraping.
- **BeautifulSoup** : Bibliothèque Python utilisée pour l'analyse syntaxique (parsing) du HTML des pages web.
- **Selenium** : Outil d'automatisation de navigateur web, indispensable pour interagir avec les pages web dynamiques de Google Maps (e.g., défilement pour charger plus d'avis, gestion des pop-ups).

Le processus implique l'identification des URL des agences bancaires marocaines, la navigation sur ces pages, la simulation des interactions utilisateur pour charger tous les avis, et l'extraction structurée des informations clés : texte de l'avis, note (étoiles), date de publication, et nom de l'agence/banque.

## 3.2 Champs Collectés par le Scraping

La phase de collecte des données est fondamentale, car elle détermine la richesse et la pertinence des analyses ultérieures. Lors de notre processus de scraping sur Google Maps, nous nous sommes concentrés sur l'extraction de plusieurs champs clés pour chaque avis client, tels que visualisés dans notre fichier `raw_data.json`. Ces informations, une fois structurées, constituent la matière première de notre pipeline de données.

Voici les principaux champs que nous avons scrappés, directement tirés de la structure JSON de nos données brutes :

- **bank\_name (Nom de la Banque) :** Il s'agit du nom de l'établissement bancaire ou de la marque à laquelle l'agence appartient (par exemple, "Attijariwafa Bank"). Ce champ est essentiel pour l'agrégation et l'analyse des avis par entité bancaire.
- **bank\_rating (Note Globale de la Banque) :** Ce champ semble indiquer la note globale moyenne de l'agence ou de la banque, bien qu'il contienne ici des valeurs répétitives ("2,2"). Il pourrait potentiellement être utilisé pour contextualiser les avis individuels par rapport à une évaluation agrégée.
- **bank\_location (Localisation de l'Agence) :** L'adresse physique de l'agence bancaire concernée par l'avis (par exemple, "4 Bd Mohamed Zerktouni, Marrakech 40000"). Cette information est cruciale pour les analyses géospatiales ou pour identifier les performances par agence spécifique.
- **review\_date (Date de l'Avis) :** La date relative à laquelle l'avis a été publié par le client (par exemple, "il y a 2 mois", "il y a 4 mois", "il y a un an"). Bien que relative dans les données brutes, cette information est fondamentale pour l'analyse des tendances temporelles.
- **review\_text (Texte de l'Avis) :** Le contenu textuel de l'avis laissé par le client (par exemple, "Une agence très mauvaise, et les employés ne respectent pas leurs clients."). C'est le champ principal sur lequel seront appliquées les techniques de Traitement du Langage Naturel (NLP) telles que l'analyse de sentiment et l'extraction de thèmes.
- **review\_rating (Note de l'Avis) :** La note spécifique attribuée par le client à cet avis particulier, souvent exprimée en étoiles ou sur une échelle numérique (par exemple, "1"). Ce champ fournit une mesure quantitative directe de la satisfaction individuelle.

Ces champs sont systématiquement extraits pour chaque avis trouvé et stockés au format JSON (comme illustré dans `raw_data.json`), ce qui facilite leur intégration et leur traitement dans les étapes ultérieures de notre pipeline de données.

```
[
  {
    "bank_name": "Attijariwafa Bank",
    "bank_rating": "2,2",
    "bank_location": "4 Bd Mohamed Zerktouni, Marrakech 40000",
    "review_date": "il y a 2 mois",
    "review_text": "Une agence très mauvaise, et les employés ne respectent pas leurs clients. Quant au responsable de l'agence,
    "review_rating": 1
  },
  {
    "bank_name": "Attijariwafa Bank",
    "bank_rating": "2,2",
    "bank_location": "4 Bd Mohamed Zerktouni, Marrakech 40000",
    "review_date": "il y a 4 mois",
    "review_text": "Je tiens à exprimer ma déception concernant cette agence. À chaque fois que je me rends sur place pour obten
    "review_rating": 1
  },
]
```

Figure 3: Champs Collectés par le Scraping

### 3.3 Automatisation

L'ensemble du processus de collecte est orchestré par Apache Airflow. Des DAGs (Directed Acyclic Graphs) sont configurés pour planifier et exécuter régulièrement les scripts de scraping, assurant ainsi une mise à jour continue des données.

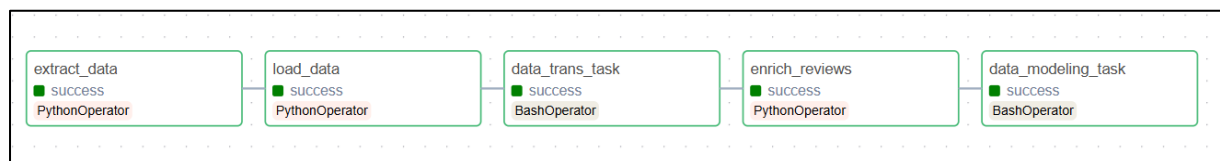


Figure 4 : Airflow Graph

### 3.4 Défis rencontrés et solutions

Scrapper les avis de Google Maps fut un véritable parcours d'obstacles. Les principales difficultés provenaient de l'absence d'une API publique dédiée, de la détection sophistiquée des bots par Google, et de la structure dynamique des pages, rendant l'extraction directe complexe. Pour contourner ces défis, nous avons adopté une approche hybride de scraping. En combinant la robustesse de BeautifulSoup pour l'analyse HTML et la capacité d'interaction de Selenium avec les éléments dynamiques, nous avons simulé un comportement utilisateur réaliste. Des stratégies complémentaires, comme la rotation d'adresses IP et la gestion des délais aléatoires, ont permis de minimiser les blocages et d'assurer une collecte de données efficace et discrète.

## 4 Phase 2 : Nettoyage et Transformation des Données

Cette phase constitue un maillon essentiel du pipeline de données. Elle vise à assurer la qualité, la cohérence et la richesse sémantique des données avant leur modélisation et leur analyse. Elle a été réalisée principalement à l'aide de DBT (Data Build Tool), associé à des requêtes SQL.

### 4.1 Nettoyage des Données

Dans un premier temps, les données brutes issues du scraping ont été nettoyées. Les opérations suivantes ont été appliquées :

- **Suppression des doublons** : Les avis identiques ou répétés ont été éliminés pour éviter les biais dans les analyses statistiques.
- **Normalisation du texte** : Le texte des avis a été transformé en minuscules, la ponctuation a été supprimée, et les mots vides (stop words) ont été éliminés pour préparer les données aux tâches de traitement du langage.
- **Traitement des valeurs manquantes** : Les enregistrements incomplets ont été soit supprimés, soit complétés par des valeurs par défaut selon les cas.

### 4.2 Enrichissement des Données

Une fois les données nettoyées, elles ont été enrichies avec des dimensions supplémentaires permettant des analyses plus approfondies :

- **Détection de la langue** : L'algorithme a identifié la langue de chaque avis pour permettre une analyse multilingue.
- **Analyse de sentiment** : Chaque avis a été classé en trois catégories : *positif*, *négatif* ou *neutre*, à l'aide d'un modèle d'analyse de sentiment pré-entraîné.
- **Extraction de thématiques** : Un algorithme de type LDA (Latent Dirichlet Allocation) a permis d'identifier les sujets fréquemment abordés dans les avis (par exemple : file d'attente, accueil, disponibilité des guichets).

id	bank_name big	bank_rating	bank_location	review_date	review_text	
1	0	Attijariwafa Ba...	2.2	4 Bd Mohamed Zerkouni, Marrakech 40000	il y a 2 mois	Une agence très mauvaise, et les employés ne respectent pas leurs clients. Quant au responsable de l'agence, il manque de professionnalisme : c'est un enfant qui
2	1	Attijariwafa Ba...	2.2	4 Bd Mohamed Zerkouni, Marrakech 40000	il y a 4 mois	Je tiens à exprimer ma déception concernant cette agence. À chaque fois que je me rends sur place pour obtenir des renseignements, il est extrêmement difficile c
3	2	Attijariwafa Ba...	2.2	4 Bd Mohamed Zerkouni, Marrakech 40000	il y a 5 mois	Je vous déconseille fortement de retirer de l'argent de ce guichet ou de collaborer avec cette agence, j'ai mis ma carte pour retirer la somme de 700 dhs ils ont été
4	3	Attijariwafa Ba...	2.2	4 Bd Mohamed Zerkouni, Marrakech 40000	il y a un an	Service absent: Ne pas tenir compte des horaires d'ouverture. Personne n'est en poste. Attente minimum une heure
5	4	Attijariwafa Ba...	2.2	4 Bd Mohamed Zerkouni, Marrakech 40000	il y a 4 mois	Le mec au guichet a refusé me faire virement déplacé, il voulait que j'aille a mon agence à Oujda, ils sont fous c'est quoi ce service
6	5	Attijariwafa Ba...	2.2	4 Bd Mohamed Zerkouni, Marrakech 40000	il y a 6 mois	Aucun sens du service... Donne l'impression de les dérangé... Je regrette d'avoir ouvert mon compte là-bas... Si vous comptez ouvrir un compte, rebroussez le che
7	6	Attijariwafa Ba...	2.2	4 Bd Mohamed Zerkouni, Marrakech 40000	il y a 7 mois	Service trop long le personnel prend des appels personnel avant de vous servir partent pendant de longue longue minute on se croirait au consulat le service télé
8	7	Attijariwafa Ba...	2.2	4 Bd Mohamed Zerkouni, Marrakech 40000	il y a un an	Mon expérience avec cette banque a été extrêmement décevante en raison de leur manque de professionnalisme flagrant. Après avoir soumis ma demande de crédi
9	8	Attijariwafa Ba...	2.2	4 Bd Mohamed Zerkouni, Marrakech 40000	il y a 2 ans	Honteux! je me suis rendu pour changer de l'argent à leur distributeur le soir (la banque étant fermée) j'insère 110€ j'attends c'est écrit traitement en cours puis l
10	9	Attijariwafa Ba...	2.2	4 Bd Mohamed Zerkouni, Marrakech 40000	il y a un an	Bad service they don't answer calls and they disrespect their clients ashamed to have an agency like that
11	9	Attijariwafa Ba...	2.2	4 Bd Mohamed Zerkouni, Marrakech 40000	il y a 2 ans	Une agence faite d'un personnel totalement incompétent qui ne font qu'à leur tâches! Sortent et ne reviennent plus! Personne ne répond un téléphone!
12	10	Attijariwafa Ba...	2.2	4 Bd Mohamed Zerkouni, Marrakech 40000	il y a 5 ans	Extrêmement lents et inefficaces, le type dans le bureau passe de longs coups de fil personnel alors qu'il y a des clients qui attendent, et fait passer ses copains a
13	11	Attijariwafa Ba...	2.2	4 Bd Mohamed Zerkouni, Marrakech 40000	il y a 4 ans	L'accueil est nul... le personnel peu qualifié... après 7 longues années... rien ne change la bas... je quitte.
14	12	Attijariwafa Ba...	2.2	4 Bd Mohamed Zerkouni, Marrakech 40000	il y a 3 ans	Avance avec l'ancien directeur c'est bien géré.
15	13	Attijariwafa Ba...	2.2	4 Bd Mohamed Zerkouni, Marrakech 40000	il y a 4 ans	Personnel incompétent! Il faut à mon avis faire le ménage et restructurer cette agence!
16	14	Attijariwafa Ba...	2.2	4 Bd Mohamed Zerkouni, Marrakech 40000	il y a 6 ans	Une agence bien situé près de l'hôtel al kabir. les agents y sont très accueillant et aimable.
17	15	Attijariwafa Ba...	2.2	4 Bd Mohamed Zerkouni, Marrakech 40000	il y a 5 ans	Service hyper nul, conseillers inspectueux.
18	16	Attijariwafa Ba...	2.2	4 Bd Mohamed Zerkouni, Marrakech 40000	il y a 7 ans	Le conseiller entreprise est complètement débordé... a évite!
19	17	Attijariwafa Ba...	2.2	4 Bd Mohamed Zerkouni, Marrakech 40000	il y a 4 ans	Aucun respect
20	18	Attijariwafa Ba...	2.2	4 Bd Mohamed Zerkouni, Marrakech 40000	il y a 3 ans	Téléphone toujours sans réponse
21	19	Attijariwafa Ba...	2.2	4 Bd Mohamed Zerkouni, Marrakech 40000	il y a 7 ans	fraie bancaire très élevé.
22	20	Attijariwafa Ba...	2.2	4 Bd Mohamed Zerkouni, Marrakech 40000	il y a 2 ans	Téléphone personne répond
23	21	Attijariwafa Ba...	2.2	4 Bd Mohamed Zerkouni, Marrakech 40000	il y a 2 ans	Agence à éviter
24	22	Attijariwafa Ba...	2.2	4 Bd Mohamed Zerkouni, Marrakech 40000	il y a 2 ans	Service inefficace
25	23	Attijariwafa Ba...	2.2	4 Bd Mohamed Zerkouni, Marrakech 40000	il y a 6 ans	Bonne continuation
26	24	Attijariwafa Ba...	2.2	4 Bd Mohamed Zerkouni, Marrakech 40000	il y a 2 ans	Mauvaise service
27	25	Attijariwafa Ba...	2.2	4 Bd Mohamed Zerkouni, Marrakech 40000	il y a 11 mois	This building is run by absolute children. It's an embarrassment to Moroccans everywhere.

Figure 5: Données Brutes

id	bank_name	bank_rating	bank_location	branch_name	review_date	en_review
1	0	Attijariwafa Ba...	2.2	4 bd mohamed zerkouni, marrakech 40000	2025-03-15	bad agency employees respect customers agency manager lacks professionalism child spends time
2	1	Attijariwafa Ba...	2.2	4 bd mohamed zerkouni, marrakech 40000	2025-01-15	want express disappointment concerning agency whenever go get information extremely difficult find
3	2	Attijariwafa Ba...	2.2	4 bd mohamed zerkouni, marrakech 40000	2024-12-15	strongly advise withdraw money window collaborate agency put card withdraw sum dhs extinguished
4	3	Attijariwafa Ba...	2.2	4 bd mohamed zerkouni, marrakech 40000	2024-05-15	absent service take account opening hours one office minimum wait one hour
5	4	Attijariwafa Ba...	2.2	4 bd mohamed zerkouni, marrakech 40000	2025-01-15	guy counter refused make move wanted go agency ouajda crazy service
6	5	Attijariwafa Ba...	2.2	4 bd mohamed zerkouni, marrakech 40000	2024-09-15	sense service give impression disturbing regret opened account plan open account turn path better n
7	6	Attijariwafa Ba...	2.2	4 bd mohamed zerkouni, marrakech 40000	2024-10-15	service long staff take personal calls serving leave long minute long minute feels like consulate plifu
8	7	Attijariwafa Ba...	2.2	4 bd mohamed zerkouni, marrakech 40000	2024-05-15	experience bank extremely disappointing due lack flagrant professionalism submitting credit request
9	8	Attijariwafa Ba...	2.2	4 bd mohamed zerkouni, marrakech 40000	2023-05-15	ashamed went change money distributor evening bank closed insert await written treatment progress
10	9	Attijariwafa Ba...	2.2	4 bd mohamed zerkouni, marrakech 40000	2023-05-15	agency made totally incompetent staff heads come come back one answers phone put check days la
11	10	Attijariwafa Ba...	2.2	4 bd mohamed zerkouni, marrakech 40000	2020-05-15	extremely slow ineffective guy office spends long personal phone calls customers waiting puts friend
12	11	Attijariwafa Ba...	2.2	4 bd mohamed zerkouni, marrakech 40000	2021-05-15	welcome zero low skilled staff long years nothing changes leave
13	12	Attijariwafa Ba...	2.2	4 bd mohamed zerkouni, marrakech 40000	2022-05-15	advance former director well managed new director salesperson agency disaster
14	13	Attijariwafa Ba...	2.2	4 bd mohamed zerkouni, marrakech 40000	2021-05-15	incompetent staff opinion necessary clean restructure agency
15	14	Attijariwafa Ba...	2.2	4 bd mohamed zerkouni, marrakech 40000	2019-05-15	well located agency near al kabir hotel agents welcoming kind
16	15	Attijariwafa Ba...	2.2	4 bd mohamed zerkouni, marrakech 40000	2020-05-15	hyper zero service disrespectful advisers
17	16	Attijariwafa Ba...	2.2	4 bd mohamed zerkouni, marrakech 40000	2018-05-15	company advisor completely overwhelmed avoids
18	17	Attijariwafa Ba...	2.2	4 bd mohamed zerkouni, marrakech 40000	2021-05-15	respect usual service
19	18	Attijariwafa Ba...	2.2	4 bd mohamed zerkouni, marrakech 40000	2022-05-15	telephone always unanswered
20	19	Attijariwafa Ba...	2.2	4 bd mohamed zerkouni, marrakech 40000	2018-05-15	high bank fees
21	20	Attijariwafa Ba...	2.2	4 bd mohamed zerkouni, marrakech 40000	2023-05-15	phone person answers
22	21	Attijariwafa Ba...	2.2	4 bd mohamed zerkouni, marrakech 40000	2023-05-15	agency avoid
23	22	Attijariwafa Ba...	2.2	4 bd mohamed zerkouni, marrakech 40000	2023-05-15	service inefficace
24	23	Attijariwafa Ba...	2.2	4 bd mohamed zerkouni, marrakech 40000	2017-05-15	best wishes
25	24	Attijariwafa Ba...	2.2	4 bd mohamed zerkouni, marrakech 40000	2023-05-15	mauvaise service
26	25	Attijariwafa Ba...	2.2	4 bd mohamed zerkouni, marrakech 40000	2024-06-15	building run absolute children embarrassment moroccans everywhere idea anyone would business pc
27	26	Attijariwafa Ba...	2.2	4 bd mohamed zerkouni, marrakech 40000	2024-05-15	bad service never met lifebranch manager level speaking dogs eating marrakech plaza branch

Figure 6 : Données transformées

## 5 Phase 3 : Modélisation des Données

Cette phase vise à structurer les données nettoyées et enrichies dans une architecture relationnelle optimisée pour l'analyse, à travers un schéma en étoile (Star Schema).

### 5.1 Conception du Data Mart

La modélisation a été réalisée dans PostgreSQL, en suivant les principes d'un schéma en étoile, avec :

- **Table de faits** : fact\_reviews, contenant les avis des clients ainsi que les indicateurs analytiques (sentiment, date, etc.).
- **Tables de dimensions** :
  - dim\_bank : Informations sur la banque (nom, code, etc.).
  - dim\_branch : Données sur l'agence bancaire (identifiant, nom, type de service, etc.).
  - dim\_location : Détails géographiques (ville, région, coordonnées GPS).
  - dim\_sentiment : Catégorisation et scores de sentiment associés aux avis.

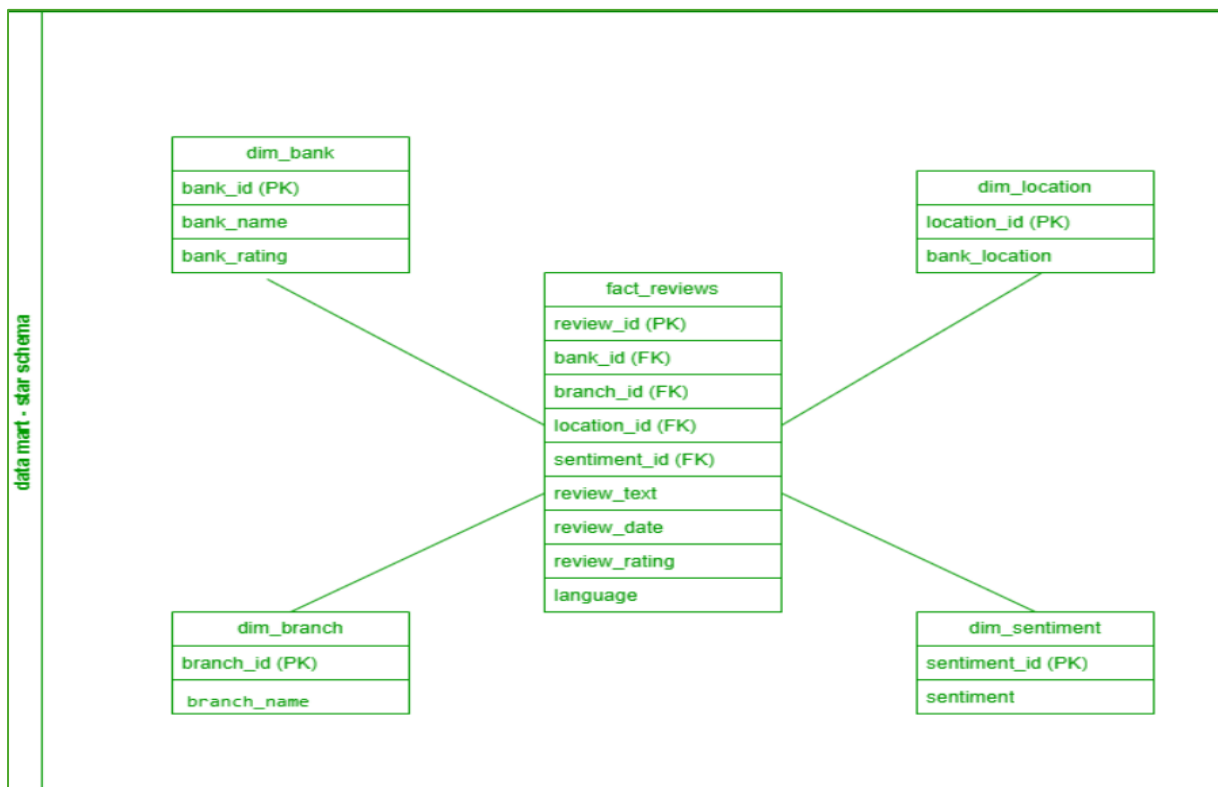


Figure 7 : Schéma ER du modèle en étoile

## 5.2 Chargement et Orchestration

Les transformations ont été implémentées via DBT, ce qui a permis de structurer les modèles de données en différentes couches logiques (staging, intermediate, marts). Le chargement des données a été automatisé à l'aide d'Apache Airflow, assurant la reproductibilité et la fiabilité du pipeline.

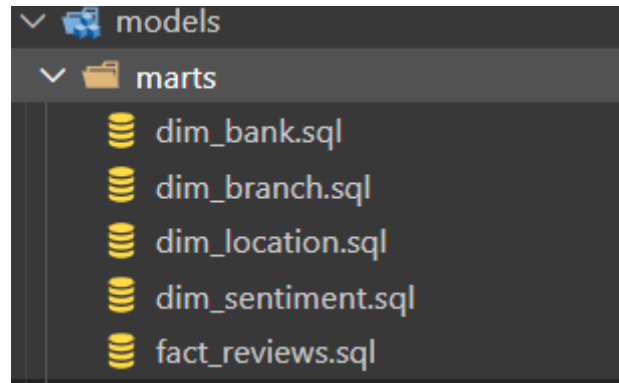


Figure 8 : Modèle DBT de transformation



## 6 Phase 4 : Analyse des Données et Reporting

Une fois les données nettoyées, enrichies et structurées dans un modèle en étoile, la dernière phase du projet a consisté à les analyser et à en extraire des indicateurs visuels exploitables via des tableaux de bord interactifs conçus avec Looker Studio.

### 6.1 Connexion des Données via Aiven

Pour exposer la base de données PostgreSQL au moteur de visualisation, nous avons utilisé Aiven, une plateforme cloud de gestion de bases de données. Elle a permis :

- La mise en place rapide et sécurisée d'une instance PostgreSQL hébergée dans le cloud.
- La configuration d'un accès public ou restreint à la base, afin de permettre la connexion à distance via Looker Studio.
- La gestion simplifiée des paramètres de sécurité, d'authentification et de haute disponibilité.

Cette solution a facilité l'intégration des données dans Looker Studio sans avoir à maintenir une infrastructure locale, tout en assurant des performances fiables pour les requêtes d'analyse.

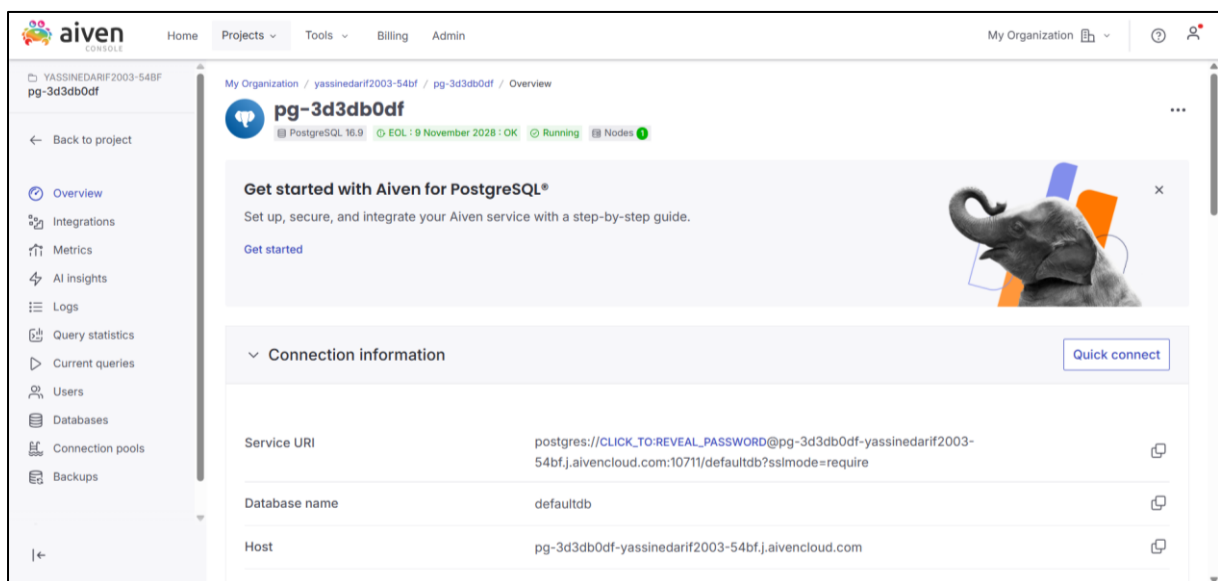


Figure 9 : Connexion des Données via Aiven

### 6.2 Construction de Tableaux de Bord Interactifs

L'outil Looker Studio (anciennement Google Data Studio) a été choisi pour sa facilité d'intégration avec PostgreSQL via Aiven, ainsi que pour sa capacité à produire des

visualisations dynamiques et personnalisables. Plusieurs tableaux de bord ont été développés, ciblant différentes dimensions d’analyse :

- **Tendance des sentiments par banque et par agence** : Ce graphique permet de visualiser l’évolution des avis positifs, négatifs et neutres au fil du temps pour chaque institution et ses agences.

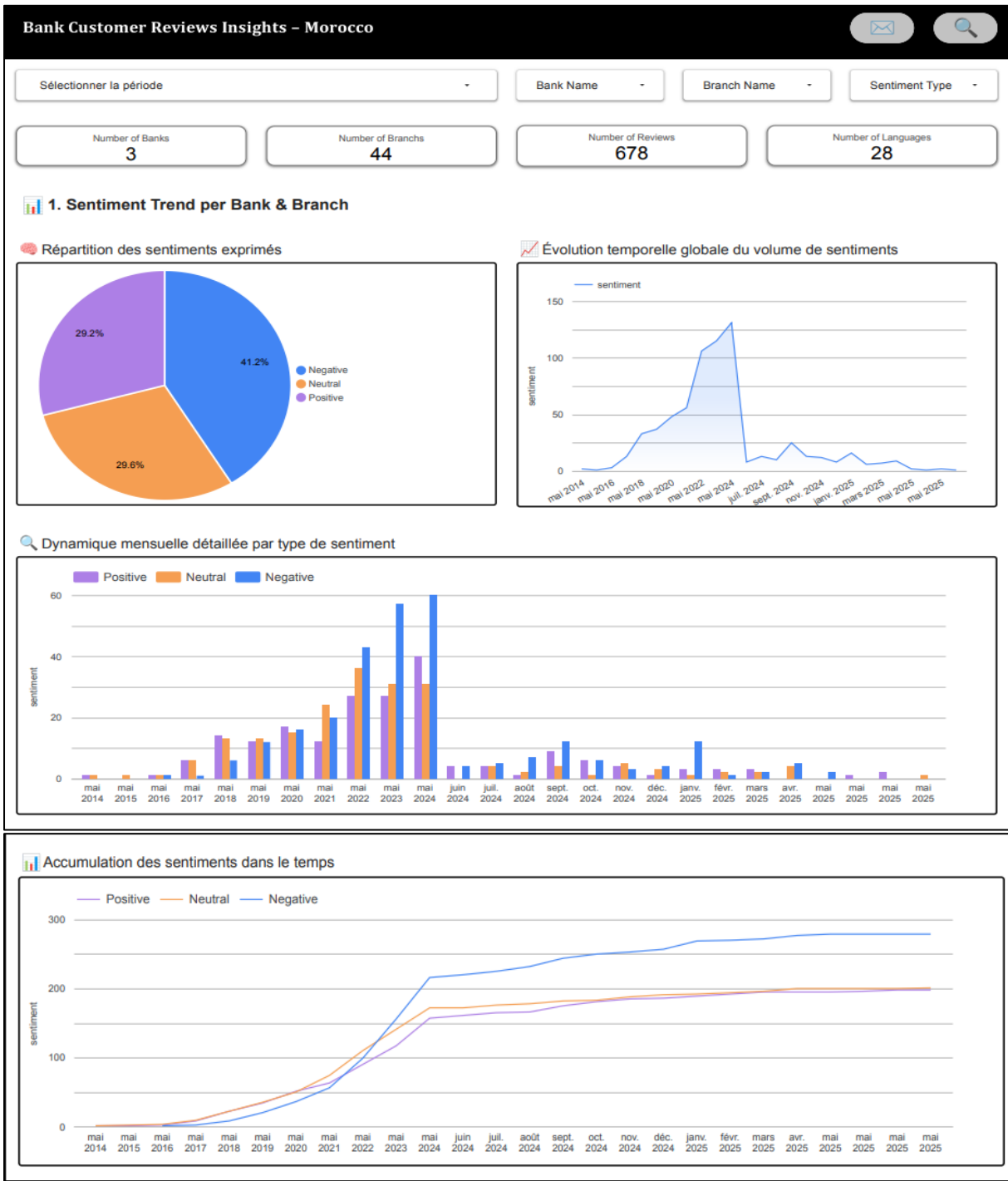


Figure 10 : Tendance des sentiments par banque et par agence

- **Sujets les plus positifs et les plus négatifs** : Une analyse thématique des avis permet d'identifier les points forts et les points faibles fréquemment évoqués par les clients.

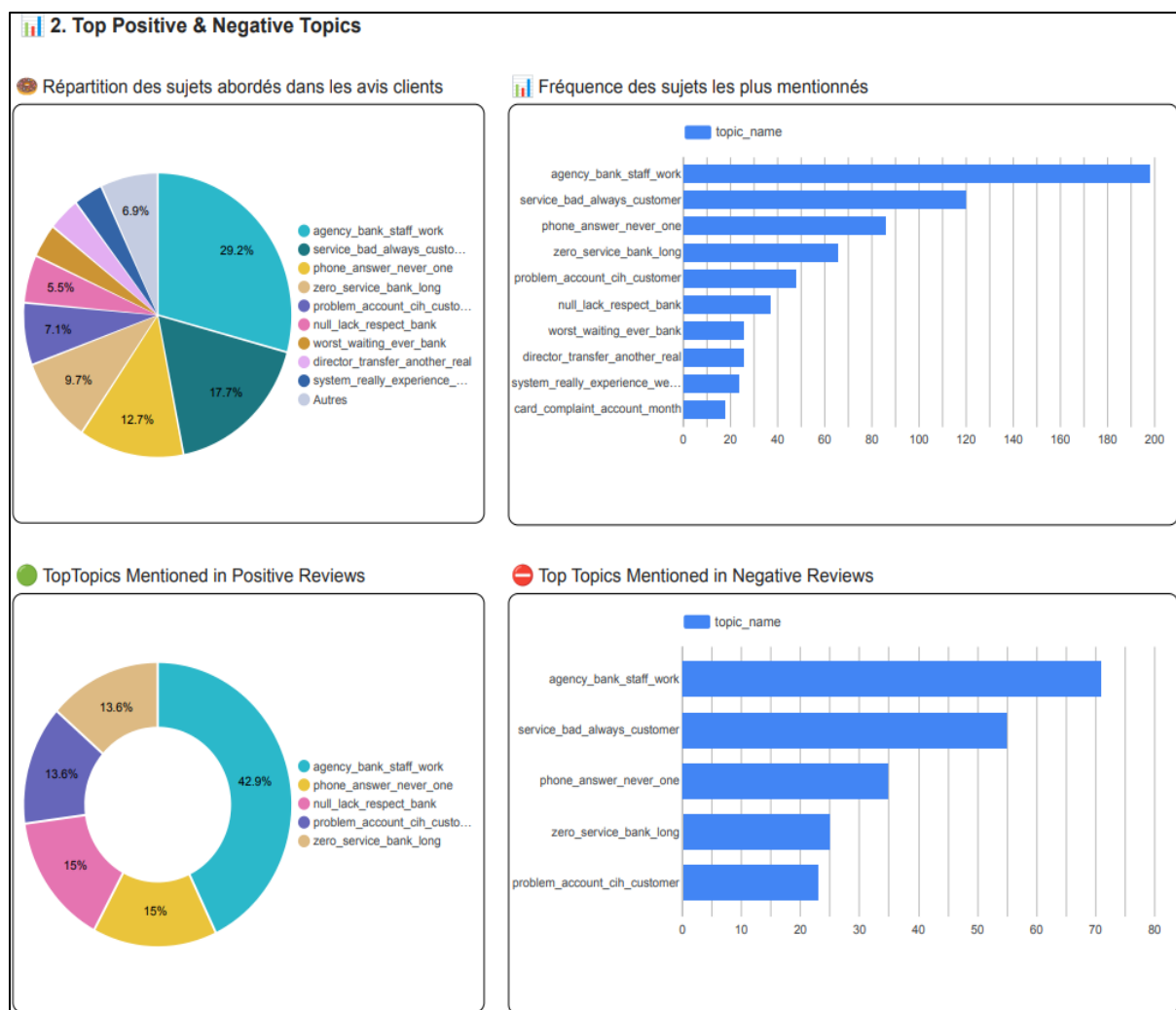


Figure 11 : Sujets les plus positifs et les plus négatifs

- **Classement de performance des agences** : Un score global de satisfaction a été calculé pour chaque agence, tenant compte du volume et de la polarité des avis. Les agences ont ensuite été classées pour repérer les meilleures et les moins performantes.

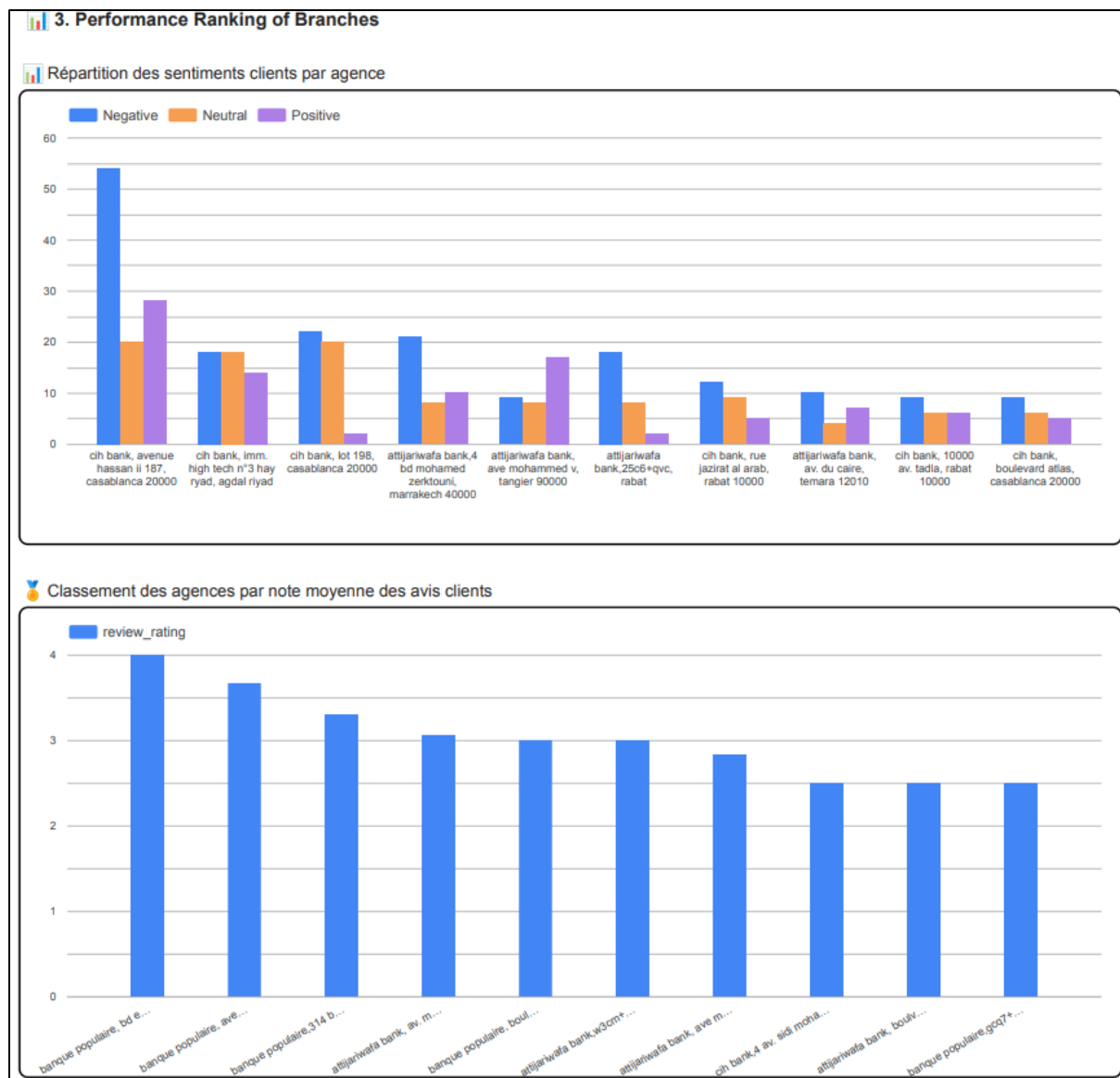


Figure 12 : Classement de performance des agences

- **Indicateurs d'expérience client** : Des KPI synthétiques ont été mis en évidence pour guider les décisions d'amélioration.

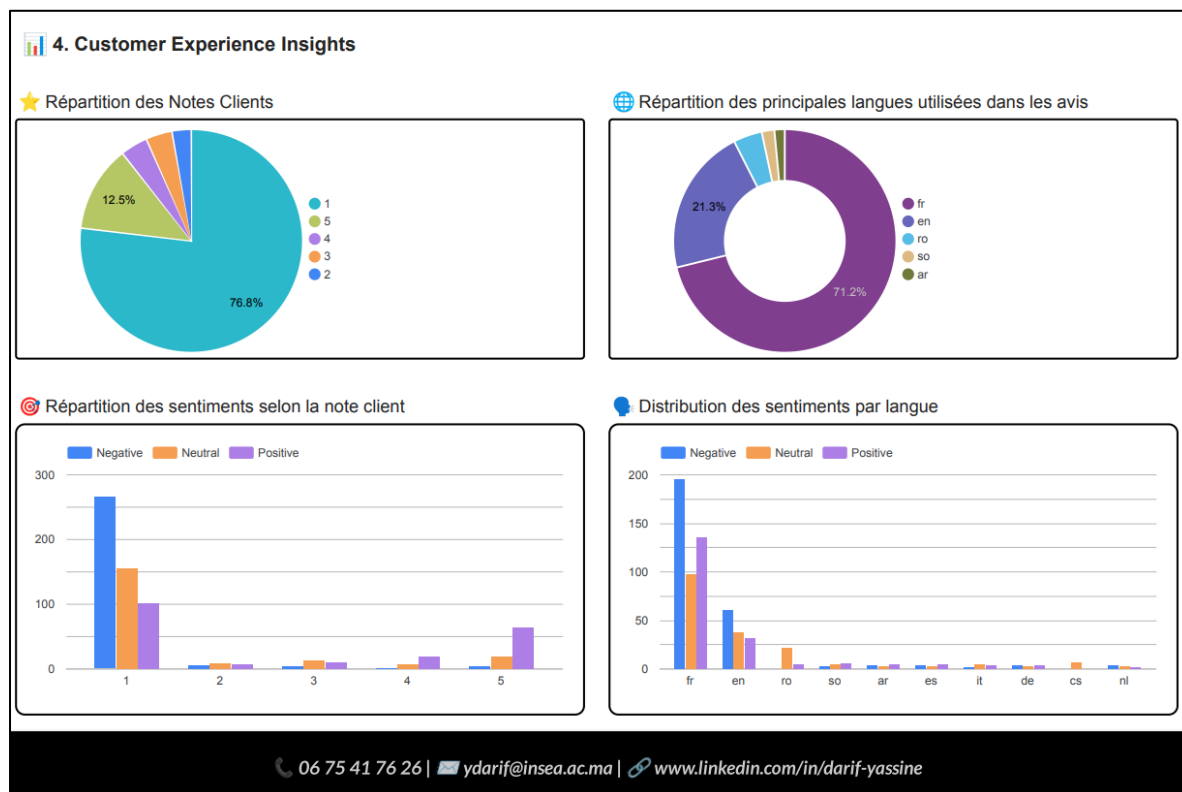


Figure 13 : Indicateurs d'expérience client

## 7 Stack Technologique Détaillée

Le choix des outils pour ce projet a été guidé par les principes d'une "Modern Data Stack" : modularité, scalabilité, facilité d'intégration et communauté active.

Étape	Outil(s)	Description
<b>Collecte</b>	Python, BeautifulSoup, Selenium	<b>Python</b> : Langage de programmation polyvalent. <b>BeautifulSoup</b> : Parsing HTML/XML pour l'extraction de données. <b>Selenium</b> : Automatisation du navigateur pour collecter les contenus dynamiques.
<b>Orchestration</b>	Apache Airflow	Plateforme open-source permettant d'automatiser, planifier et surveiller des workflows complexes sous forme de DAGs (Directed Acyclic Graphs).
<b>Stockage</b>	PostgreSQL	Système de gestion de base de données relationnelle open-source, robuste et adapté au stockage structuré des données brutes et transformées.
<b>Transformation</b>	DBT (Data Build Tool)	Outil qui permet aux ingénieurs data de transformer les données dans un entrepôt de données, en utilisant uniquement SQL. Il offre le versionnement, les tests et la documentation.
<b>Analyse &amp; BI</b>	Looker Studio	Outil de Business Intelligence de Google. Il permet de créer des tableaux de bord et des rapports à partir de différentes sources de données.
<b>Contrôle de version</b>	GitHub	Plateforme basée sur Git, essentielle pour la collaboration, le versionnement du code source, les scripts SQL, les DAGs, et la gestion des déploiements.

## 8 Perspectives d'Amélioration

Ce projet pose des bases solides, mais plusieurs pistes d'amélioration peuvent être envisagées pour enrichir davantage l'analyse et la valeur ajoutée :

- **Ajout de données internes bancaires** : Intégrer les avis collectés avec des données internes des banques (e.g., données CRM, historiques de transactions, données sur les incidents, informations sur le personnel) permettrait une analyse croisée plus poussée et l'identification de corrélations entre les retours clients et les opérations internes.
- **Résumé automatique des avis avec l'IA** : Développer ou intégrer des modèles de résumé textuel (e.g., basés sur des architectures Transformer comme BERT ou GPT) pour générer des synthèses concises des avis, notamment pour les avis négatifs, afin d'accélérer la compréhension des problèmes par les équipes opérationnelles.
- **Analyse de la satisfaction par type de service/produit** : Affiner l'extraction de thèmes pour lier spécifiquement les avis à des services ou produits bancaires précis (e.g., crédit immobilier, carte bancaire, compte courant, application mobile), offrant ainsi une granularité d'analyse supérieure.
- **Détection d'anomalies et alertes en temps réel** : Implémenter des mécanismes pour détecter des changements abrupts dans le sentiment client ou l'apparition de nouveaux thèmes négatifs, et générer des alertes automatiques pour une intervention rapide des banques.
- **Extension aux plateformes sociales** : Collecter des avis et des discussions sur d'autres plateformes sociales (Facebook, Twitter) pour avoir une vue plus complète de la "voix du client".
- **Amélioration des modèles NLP pour le dialecte marocain** : L'arabe marocain (Darija) a des spécificités. L'entraînement de modèles NLP spécifiques à ce dialecte pourrait considérablement améliorer la précision de l'analyse de sentiments et de la modélisation thématique.
- **Prédiction du taux de désabonnement (Churn)** : Corréler les sentiments négatifs avec des indicateurs de désabonnement potentiel pour anticiper les départs de clients.

## 9 Conclusion

Ce projet a démontré la valeur considérable de l'exploitation des données clients non structurés, spécifiquement les avis Google Maps, pour le secteur bancaire marocain. En mettant en place une chaîne de traitement de données complète et moderne (Modern Data Stack), de la collecte à la visualisation, nous avons réussi à transformer des informations brutes en insights actionnables.

La capacité à analyser les sentiments, à identifier les thèmes dominants et à évaluer la performance des agences offre aux banques un outil puissant pour comprendre la perception de leurs services par leurs clients. Cette chaîne de traitement est non seulement fonctionnelle mais également répliquable et scalable, permettant d'intégrer de nouvelles sources de données ou d'étendre l'analyse à d'autres régions ou secteurs.

Les résultats obtenus confirment que l'investissement dans l'analyse des retours clients digitaux est essentiel pour améliorer l'expérience client, renforcer la satisfaction et maintenir un avantage concurrentiel dans un marché de plus en plus exigeant. Ce projet ouvre la voie à des analyses plus sophistiquées et à une prise de décision basée sur les données pour l'amélioration continue des services bancaires.