

# **PROJET D'ANALYSE DE SENTIMENT DES COMMENTAIRES BANCAIRES**

**RÉALISÉ PAR :**  
**HICHAM FADLI**  
**EZZAGRANI HABIBA**

**SUPERVISÉ PAR:**  
**M.BENELALLAM**

## Table des matières

|          |  |           |
|----------|--|-----------|
| <b>1</b> | <b>Introduction</b>  | <b>3</b>  |
| <b>2</b> | <b>Méthodologie</b>  | <b>3</b>  |
| 2.1      | Collecte des données des commentaires clients . . . . .  | 3         |
| 2.2      | Prétraitement des données . . . . .  | 5         |
| 2.3      | Utilisation du modèle BERT pour l'analyse de sentiment . . . . .                               | 6         |
| 2.4      | Chargement des données dans le datawarehouse . . . . .   | 7         |
| <b>3</b> | <b>Résultats et recommandations</b>  | <b>11</b> |
| <b>4</b> | <b>Limites des modèles pré-entraînés et de l'analyse de sentiment automatisée</b>              | <b>14</b> |
| <b>5</b> | <b>Outils utilisés</b>   | <b>15</b> |
| <b>6</b> | <b>Conclusion</b>  | <b>17</b> |
| <b>7</b> | <b>Annexes</b>   | <b>19</b> |
| 7.1      | Code source utilisé pour effectuer un scrapping de données à partir de l'API d'Apify . . . . . | 19        |
| 7.2      | Code source utilisé pour filtrer les caractères spéciaux : . . . . .                           | 19        |
| 7.3      | Code source utilisé pour l'analyse de sentiment : . . . . .                                    | 20        |
| 7.4      | code source de DAG (Directed Acyclic Graph) pour définir le flux de travail Airflow . . . . .  | 21        |
| <b>8</b> | <b>Conclusion générale</b>   | <b>22</b> |

## Liste des figures

|    |   |    |
|----|---|----|
| 1  | Configuration des tâches de scraping avec Apify . . . . .         | 4  |
| 2  | Exécution des tâches de scraping . . . . .                        | 4  |
| 3  | Nettoyage et prétraitement des données . . . . .                  | 5  |
| 4  | prétraitement des données . . . . .                               | 6  |
| 5  | analyse de sentiment . . . . .                                    | 7  |
| 6  | schéma de données . . . . .                                       | 8  |
| 7  | Préparation de l'environnement . . . . .                          | 9  |
| 8  | Création d'une connexion entre Airflow et PostgreSQL . . . . .    | 9  |
| 9  | Création d'un DAG (Directed Acyclic Graph) avec Airflow . . . . . | 10 |
| 10 | Stockage des résultats dans PostgreSQL . . . . .                  | 10 |
| 11 | Diagramme 1 . . . . .   | 12 |
| 12 | Diagramme 2 . . . . .   | 12 |
| 13 | Diagramme 3 . . . . .   | 13 |
| 14 | Visualisation sur map . . . . .                                   | 13 |
| 15 | Apify . . . . .   | 15 |
| 16 | Python . . . . .  | 15 |
| 17 | Hugging Face . . . . .  | 16 |
| 18 | Apache Airflow . . . . .  | 16 |
| 19 | Postgresql . . . . .  | 17 |

# 1 Introduction

Le présent rapport décrit le projet d'analyse de sentiment des commentaires bancaires. Dans un environnement bancaire en constante évolution, il est essentiel pour les institutions financières de comprendre les sentiments et les opinions des clients exprimés dans leurs commentaires. L'analyse de sentiment des commentaires bancaires permet d'extraire des informations précieuses à partir de ces données non structurées et de prendre des décisions éclairées pour améliorer l'expérience client et optimiser les services offerts.

L'objectif principal de ce projet est d'appliquer des techniques de traitement du langage naturel (NLP) pour analyser le sentiment des commentaires bancaires. Plus précisément, nous visons à déterminer si les commentaires des clients sont positifs, négatifs ou neutres. Cette analyse permettra aux banques de mieux comprendre la perception de leurs clients et de détecter rapidement les problèmes potentiels ou les domaines nécessitant des améliorations.

En utilisant des modèles de NLP pré-entraînés tels que BERT (Bidirectional Encoder Representations from Transformers) ou d'autres méthodes d'analyse de sentiment, nous pourrions extraire des informations précieuses à partir des commentaires bancaires collectés. Ces informations pourront ensuite être utilisées pour prendre des décisions stratégiques, améliorer les produits et services, et renforcer la relation avec les clients.

Le rapport suivant présentera en détail la méthodologie utilisée pour mener à bien ce projet, les résultats obtenus et les recommandations pour une utilisation optimale de l'analyse de sentiment des commentaires bancaires.

## 2 Méthodologie

### 2.1 Collecte des données des commentaires clients

Dans le cadre de notre projet d'analyse de sentiment des commentaires bancaires, nous avons utilisé l'outil Apify pour la collecte des données. Apify est une plateforme de scraping et d'automatisation qui permet de récupérer des informations à partir de sites web de manière efficace et évolutive. Voici notre approche pour la collecte des données des commentaires clients :

**Identification des sources de données pertinentes :** Nous avons identifié les plateformes de médias sociaux, les sites d'avis en ligne et les bases de données internes des banques comme principales sources de commentaires clients. Nous avons effectué une analyse approfondie pour sélectionner les sources les plus pertinentes en fonction de la nature du projet et de l'objectif d'analyse de sentiment.

**Configuration des tâches de scraping avec Apify :** À l'aide de l'interface conviviale d'Apify, nous avons configuré des tâches de scraping pour chaque source de données identifiée. Pour chaque source, nous avons défini les règles de sélection

des éléments à extraire, tels que les commentaires, les évaluations et les métadonnées associées. Nous avons également ajusté les paramètres de scraping, tels que les délais de requêtes et les stratégies de récupération, pour garantir un processus rapide et fiable.

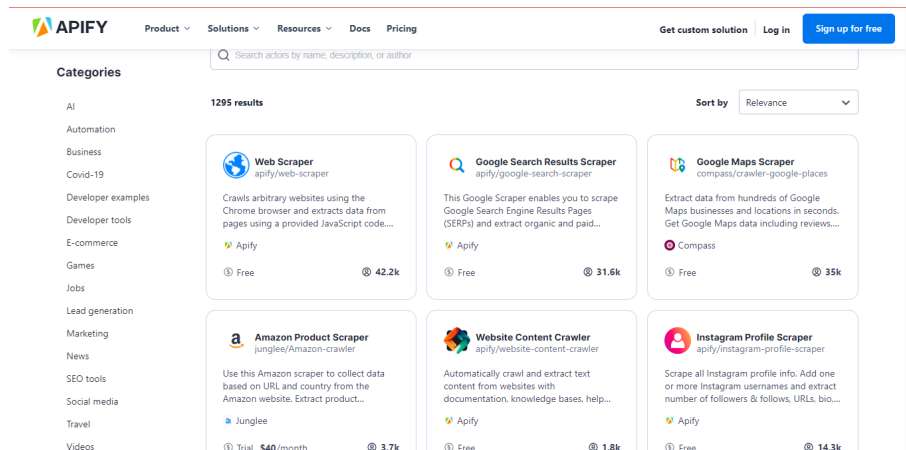


FIGURE 1 – Configuration des tâches de scraping avec Apify

**Exécution des tâches de scraping :** Une fois les tâches de scraping configurées, nous les avons lancées pour extraire les données des commentaires clients à partir des différentes sources. Apify s'est chargé de collecter les données de manière automatisée, en respectant les règles et les paramètres définis.

 The image shows an Excel spreadsheet with multiple columns of data. The columns are labeled: title, subtitle, description, price, menu, category, name, address, location, neighborhood, street, city, postalCode, state, countryCode, plusCode, website, phone, location, client, business, location, hour, and user. The data rows contain information about various banks and their services, including "Crédit Immobilier et Hôtelier (CIH Bank)", "Crédit Immobilier Et Hôtelier Agence Casa Ifriquia", and "Crédit Immobilier Et Hôtelier Agence Bd (CIH)". The data includes details like addresses, phone numbers, and website URLs.

FIGURE 2 – Exécution des tâches de scraping

**Nettoyage et prétraitement des données :** Après la collecte des données, nous avons effectué une étape de nettoyage et de prétraitement pour garantir la qualité et la cohérence des données. Cela comprend la suppression des caractères

indésirables, la correction des fautes d’orthographe, la normalisation du texte, et d’autres techniques de prétraitement couramment utilisées dans le domaine du NLP.

[illegible]

FIGURE 3 – Nettoyage et prétraitement des données

## 2.2 Prétraitement des données

Le prétraitement des données est une étape cruciale pour nettoyer et préparer les données avant de les utiliser dans notre modèle d'analyse de sentiment. Les principales étapes de prétraitement que nous avons suivies sont :

- **Nettoyage des données** : Nous avons éliminé les commentaires dupliqués et effectué une étape de nettoyage pour éliminer les caractères indésirables tels que la ponctuation, les balises HTML, les URL et les chiffres. Nous avons également converti le texte en minuscules pour normaliser le texte.
- **Tokenisation** : Nous avons divisé les commentaires en mots individuels en utilisant la tokenisation. Cela facilite le traitement et l'analyse ultérieure.
- **Suppression des stopwords** : Nous avons supprimé les stopwords, tels que "et", "le", "mais", qui n'apportent pas beaucoup d'informations pour l'analyse de sentiment.
- **Normalisation des mots** : Nous avons normalisé les mots en utilisant des techniques de lemmatisation ou de stemming pour regrouper différentes formes d'un mot en une seule forme standardisée.
- **Encodage des mots** : Nous avons représenté les mots sous forme numérique en utilisant des techniques d'encodage, telles que l'encodage one-hot ou l'encodage basé sur des embeddings.
- **Traitement des données manquantes** : Nous avons analysé et traité les données manquantes en utilisant des techniques d'imputation ou d'élimination des échantillons contenant des données manquantes.

|         |  |                     |  |                      |  |        |  |              |  |          |  |         |  |          |  |           |  |             |  |      |  |              |  |         |  |          |  |   |  |   |  |   |  |   |  |   |  |   |  |   |  |   |  |  |  |
|---------|--|---------------------|--|----------------------|--|--------|--|--------------|--|----------|--|---------|--|----------|--|-----------|--|-------------|--|------|--|--------------|--|---------|--|----------|--|---|--|---|--|---|--|---|--|---|--|---|--|---|--|---|--|--|--|
| Excel   |  | CHIReviewsTokenized |  | Rechercher (Alt + Q) |  |        |  |              |  |          |  |         |  |          |  |           |  |             |  |      |  |              |  |         |  |          |  |   |  |   |  |   |  |   |  |   |  |   |  |   |  |   |  |  |  |
| Fichier |  | Accueil             |  | Insertion            |  | Dessin |  | Mise en page |  | Formules |  | Données |  | Révision |  | Affichage |  | Automatiser |  | Aide |  | Commentaires |  | Edition |  | Partager |  |   |  |   |  |   |  |   |  |   |  |   |  |   |  |   |  |  |  |
| V       |  | U                   |  | T                    |  | S      |  | R            |  | Q        |  | P       |  | O        |  | N         |  | M           |  | L    |  | K            |  | J       |  | I        |  | H |  | G |  | F |  | E |  | D |  | C |  | B |  | A |  |  |  |
| V       |  | U                   |  | T                    |  | S      |  | R            |  | Q        |  | P       |  | O        |  | N         |  | M           |  | L    |  | K            |  | J       |  | I        |  | H |  | G |  | F |  | E |  | D |  | C |  | B |  | A |  |  |  |
| V       |  | U                   |  | T                    |  | S      |  | R            |  | Q        |  | P       |  | O        |  | N         |  | M           |  | L    |  | K            |  | J       |  | I        |  | H |  | G |  | F |  | E |  | D |  | C |  | B |  | A |  |  |  |
| V       |  | U                   |  | T                    |  | S      |  | R            |  | Q        |  | P       |  | O        |  | N         |  | M           |  | L    |  | K            |  | J       |  | I        |  | H |  | G |  | F |  | E |  | D |  | C |  | B |  | A |  |  |  |
| V       |  | U                   |  | T                    |  | S      |  | R            |  | Q        |  | P       |  | O        |  | N         |  | M           |  | L    |  | K            |  | J       |  | I        |  | H |  | G |  | F |  | E |  | D |  | C |  | B |  | A |  |  |  |
| V       |  | U                   |  | T                    |  | S      |  | R            |  | Q        |  | P       |  | O        |  | N         |  | M           |  | L    |  | K            |  | J       |  | I        |  | H |  | G |  | F |  | E |  | D |  | C |  | B |  | A |  |  |  |
| V       |  | U                   |  | T                    |  | S      |  | R            |  | Q        |  | P       |  | O        |  | N         |  | M           |  | L    |  | K            |  | J       |  | I        |  | H |  | G |  | F |  | E |  | D |  | C |  | B |  | A |  |  |  |
| V       |  | U                   |  | T                    |  | S      |  | R            |  | Q        |  | P       |  | O        |  | N         |  | M           |  | L    |  | K            |  | J       |  | I        |  | H |  | G |  | F |  | E |  | D |  | C |  | B |  | A |  |  |  |
| V       |  | U                   |  | T                    |  | S      |  | R            |  | Q        |  | P       |  | O        |  | N         |  | M           |  | L    |  | K            |  | J       |  | I        |  | H |  | G |  | F |  | E |  | D |  | C |  | B |  | A |  |  |  |
| V       |  | U                   |  | T                    |  | S      |  | R            |  | Q        |  | P       |  | O        |  | N         |  | M           |  | L    |  | K            |  | J       |  | I        |  | H |  | G |  | F |  | E |  | D |  | C |  | B |  | A |  |  |  |
| V       |  | U                   |  | T                    |  | S      |  | R            |  | Q        |  | P       |  | O        |  | N         |  | M           |  | L    |  | K            |  | J       |  | I        |  | H |  | G |  | F |  | E |  | D |  | C |  | B |  | A |  |  |  |
| V       |  | U                   |  | T                    |  | S      |  | R            |  | Q        |  | P       |  | O        |  | N         |  | M           |  | L    |  | K            |  | J       |  | I        |  | H |  | G |  | F |  | E |  | D |  | C |  | B |  | A |  |  |  |
| V       |  | U                   |  | T                    |  | S      |  | R            |  | Q        |  | P       |  | O        |  | N         |  | M           |  | L    |  | K            |  | J       |  | I        |  | H |  | G |  | F |  | E |  | D |  | C |  | B |  | A |  |  |  |
| V       |  | U                   |  | T                    |  | S      |  | R            |  | Q        |  | P       |  | O        |  | N         |  | M           |  | L    |  | K            |  | J       |  | I        |  | H |  | G |  | F |  | E |  | D |  | C |  | B |  | A |  |  |  |
| V       |  | U                   |  | T                    |  | S      |  | R            |  | Q        |  | P       |  | O        |  | N         |  | M           |  | L    |  | K            |  | J       |  | I        |  | H |  | G |  | F |  | E |  | D |  | C |  | B |  | A |  |  |  |
| V       |  | U                   |  | T                    |  | S      |  | R            |  | Q        |  | P       |  | O        |  | N         |  | M           |  | L    |  | K            |  | J       |  | I        |  | H |  | G |  | F |  | E |  | D |  | C |  | B |  | A |  |  |  |
| V       |  | U                   |  | T                    |  | S      |  | R            |  | Q        |  | P       |  | O        |  | N         |  | M           |  | L    |  | K            |  | J       |  | I        |  | H |  | G |  | F |  | E |  | D |  | C |  | B |  | A |  |  |  |
| V       |  | U                   |  | T                    |  | S      |  | R            |  | Q        |  | P       |  | O        |  | N         |  | M           |  | L    |  | K            |  | J       |  | I        |  | H |  | G |  | F |  | E |  | D |  | C |  | B |  | A |  |  |  |
| V       |  | U                   |  | T                    |  | S      |  | R            |  | Q        |  | P       |  | O        |  | N         |  | M           |  | L    |  | K            |  | J       |  | I        |  | H |  | G |  | F |  | E |  | D |  | C |  | B |  | A |  |  |  |
| V       |  | U                   |  | T                    |  | S      |  | R            |  | Q        |  | P       |  | O        |  | N         |  | M           |  | L    |  | K            |  | J       |  | I        |  | H |  | G |  | F |  | E |  | D |  | C |  | B |  | A |  |  |  |
| V       |  | U                   |  | T                    |  | S      |  | R            |  | Q        |  | P       |  | O        |  | N         |  | M           |  | L    |  | K            |  | J       |  | I        |  | H |  | G |  | F |  | E |  | D |  | C |  | B |  | A |  |  |  |
| V       |  | U                   |  | T                    |  | S      |  | R            |  | Q        |  | P       |  | O        |  | N         |  | M           |  | L    |  | K            |  | J       |  | I        |  | H |  | G |  | F |  | E |  | D |  | C |  | B |  | A |  |  |  |
| V       |  | U                   |  | T                    |  | S      |  | R            |  | Q        |  | P       |  | O        |  | N         |  | M           |  | L    |  | K            |  | J       |  | I        |  | H |  | G |  | F |  | E |  | D |  | C |  | B |  | A |  |  |  |
| V       |  | U                   |  | T                    |  | S      |  | R            |  | Q        |  | P       |  | O        |  | N         |  | M           |  | L    |  | K            |  | J       |  | I        |  | H |  | G |  | F |  | E |  | D |  | C |  | B |  | A |  |  |  |
| V       |  | U                   |  | T                    |  | S      |  | R            |  | Q        |  | P       |  | O        |  | N         |  | M           |  | L    |  | K            |  | J       |  | I        |  | H |  | G |  | F |  | E |  | D |  | C |  | B |  | A |  |  |  |
| V       |  | U                   |  | T                    |  | S      |  | R            |  | Q        |  | P       |  | O        |  | N         |  | M           |  | L    |  | K            |  | J       |  | I        |  | H |  | G |  | F |  | E |  | D |  | C |  | B |  | A |  |  |  |
| V       |  | U                   |  | T                    |  | S      |  | R            |  | Q        |  | P       |  | O        |  | N         |  | M           |  | L    |  | K            |  | J       |  | I        |  | H |  | G |  | F |  | E |  | D |  | C |  | B |  | A |  |  |  |
| V       |  | U                   |  | T                    |  | S      |  | R            |  | Q        |  | P       |  | O        |  | N         |  | M           |  | L    |  | K            |  | J       |  | I        |  | H |  | G |  | F |  | E |  | D |  | C |  | B |  | A |  |  |  |
| V       |  | U                   |  | T                    |  | S      |  | R            |  | Q        |  | P       |  | O        |  | N         |  | M           |  | L    |  | K            |  | J       |  | I        |  | H |  | G |  | F |  | E |  | D |  | C |  | B |  | A |  |  |  |
| V       |  | U                   |  | T                    |  | S      |  | R            |  | Q        |  | P       |  | O        |  | N         |  | M           |  | L    |  | K            |  | J       |  | I        |  | H |  | G |  | F |  | E |  | D |  | C |  | B |  | A |  |  |  |
| V       |  | U                   |  | T                    |  | S      |  | R            |  | Q        |  | P       |  | O        |  | N         |  | M           |  | L    |  | K            |  | J       |  | I        |  | H |  | G |  | F |  | E |  | D |  | C |  | B |  | A |  |  |  |
| V       |  | U                   |  | T                    |  | S      |  | R            |  | Q        |  | P       |  | O        |  | N         |  | M           |  | L    |  | K            |  | J       |  | I        |  | H |  | G |  | F |  | E |  | D |  | C |  | B |  | A |  |  |  |
| V       |  | U                   |  | T                    |  | S      |  | R            |  | Q        |  | P       |  | O        |  | N         |  | M           |  | L    |  | K            |  | J       |  | I        |  | H |  | G |  | F |  | E |  | D |  | C |  | B |  | A |  |  |  |
| V       |  | U                   |  | T                    |  | S      |  | R            |  | Q        |  | P       |  | O        |  | N         |  | M           |  | L    |  | K            |  | J       |  | I        |  | H |  | G |  | F |  | E |  | D |  | C |  | B |  | A |  |  |  |
| V       |  | U                   |  | T                    |  | S      |  | R            |  | Q        |  | P       |  | O        |  | N         |  | M           |  | L    |  | K            |  | J       |  | I        |  | H |  | G |  | F |  | E |  | D |  | C |  | B |  | A |  |  |  |
| V       |  | U                   |  | T                    |  | S      |  | R            |  | Q        |  | P       |  | O        |  | N         |  | M           |  | L    |  | K            |  | J       |  | I        |  | H |  | G |  | F |  | E |  | D |  | C |  | B |  | A |  |  |  |
| V       |  | U                   |  | T                    |  | S      |  | R            |  | Q        |  | P       |  | O        |  | N         |  | M           |  | L    |  | K            |  | J       |  | I        |  | H |  | G |  | F |  | E |  | D |  | C |  | B |  | A |  |  |  |
| V       |  | U                   |  | T                    |  | S      |  | R            |  | Q        |  | P       |  | O        |  | N         |  | M           |  | L    |  | K            |  | J       |  | I        |  | H |  | G |  | F |  | E |  | D |  | C |  | B |  | A |  |  |  |
| V       |  | U                   |  | T                    |  | S      |  | R            |  | Q        |  | P       |  | O        |  | N         |  | M           |  | L    |  | K            |  | J       |  | I        |  | H |  | G |  | F |  | E |  | D |  | C |  | B |  | A |  |  |  |
| V       |  | U                   |  | T                    |  | S      |  | R            |  | Q        |  | P       |  | O        |  | N         |  | M           |  | L    |  | K            |  | J       |  | I        |  | H |  | G |  | F |  | E |  | D |  | C |  | B |  | A |  |  |  |
| V       |  | U                   |  | T                    |  | S      |  | R            |  | Q        |  | P       |  | O        |  | N         |  | M           |  | L    |  | K            |  | J       |  | I        |  | H |  | G |  | F |  | E |  | D |  | C |  | B |  | A |  |  |  |
| V       |  | U                   |  | T                    |  | S      |  | R            |  | Q        |  | P       |  | O        |  | N         |  | M           |  | L    |  | K            |  | J       |  | I        |  | H |  | G |  | F |  | E |  | D |  | C |  | B |  | A |  |  |  |
| V       |  | U                   |  | T                    |  | S      |  | R            |  | Q        |  | P       |  | O        |  | N         |  | M           |  | L    |  | K            |  | J       |  | I        |  | H |  | G |  | F |  | E |  | D |  | C |  | B |  | A |  |  |  |
| V       |  | U                   |  | T                    |  | S      |  | R            |  | Q        |  | P       |  | O        |  | N         |  | M           |  | L    |  | K            |  | J       |  | I        |  | H |  | G |  | F |  | E |  | D |  | C |  | B |  | A |  |  |  |
| V       |  | U                   |  | T                    |  | S      |  | R            |  | Q        |  | P       |  | O        |  | N         |  | M           |  | L    |  | K            |  | J       |  | I        |  | H |  | G |  | F |  | E |  | D |  | C |  | B |  | A |  |  |  |
| V       |  | U                   |  | T                    |  | S      |  | R            |  | Q        |  | P       |  | O        |  | N         |  | M           |  | L    |  | K            |  | J       |  | I        |  | H |  | G |  | F |  | E |  | D |  | C |  | B |  | A |  |  |  |
| V       |  | U                   |  | T                    |  | S      |  | R            |  | Q        |  | P       |  | O        |  | N         |  | M           |  | L    |  | K            |  | J       |  | I        |  | H |  | G |  | F |  | E |  | D |  | C |  | B |  | A |  |  |  |
| V       |  | U                   |  | T                    |  | S      |  | R            |  | Q        |  | P       |  | O        |  | N         |  | M           |  | L    |  | K            |  | J       |  | I        |  | H |  | G |  | F |  | E |  | D |  | C |  | B |  | A |  |  |  |
| V       |  | U                   |  | T                    |  | S      |  | R            |  | Q        |  | P       |  | O        |  | N         |  | M           |  | L    |  | K            |  | J       |  | I        |  | H |  | G |  | F |  | E |  | D |  | C |  | B |  | A |  |  |  |
| V       |  | U                   |  | T                    |  | S      |  | R            |  | Q        |  | P       |  | O        |  | N         |  | M           |  | L    |  | K            |  | J       |  | I        |  | H |  | G |  | F |  | E |  | D |  | C |  | B |  | A |  |  |  |
| V       |  | U                   |  | T                    |  | S      |  | R            |  | Q        |  | P       |  | O        |  | N         |  | M           |  | L    |  | K            |  | J       |  | I        |  | H |  | G |  | F |  | E |  | D |  | C |  | B |  | A |  |  |  |
| V       |  | U                   |  | T                    |  | S      |  | R            |  | Q        |  | P       |  | O        |  | N         |  | M           |  | L    |  | K            |  | J       |  | I        |  | H |  | G |  | F |  | E |  | D |  | C |  | B |  | A |  |  |  |
| V       |  | U                   |  | T                    |  | S      |  | R            |  | Q        |  | P       |  | O        |  | N         |  | M           |  | L    |  | K            |  | J       |  | I        |  | H |  | G |  | F |  | E |  | D |  | C |  | B |  | A |  |  |  |
| V       |  | U                   |  | T                    |  | S      |  | R            |  | Q        |  | P       |  | O        |  | N         |  | M           |  | L    |  | K            |  | J       |  | I        |  | H |  | G |  | F |  | E |  | D |  | C |  | B |  | A |  |  |  |
| V       |  | U                   |  | T                    |  | S      |  | R            |  | Q        |  | P       |  | O        |  | N         |  | M           |  | L    |  | K            |  | J       |  | I        |  | H |  | G |  | F |  | E |  | D |  | C |  | B |  | A |  |  |  |
| V       |  | U                   |  | T                    |  | S      |  | R            |  | Q        |  | P       |  | O        |  | N         |  | M           |  | L    |  | K            |  | J       |  | I        |  | H |  | G |  | F |  | E |  | D |  | C |  | B |  | A |  |  |  |
| V       |  | U                   |  | T                    |  | S      |  | R            |  | Q        |  | P       |  | O        |  | N         |  | M           |  | L    |  | K            |  | J       |  | I        |  | H |  | G |  | F |  | E |  | D |  | C |  | B |  | A |  |  |  |
| V       |  | U                   |  | T                    |  | S      |  | R            |  | Q        |  | P       |  | O        |  | N         |  | M           |  | L    |  | K            |  | J       |  | I        |  | H |  | G |  | F |  | E |  | D |  | C |  | B |  | A |  |  |  |
| V       |  | U                   |  | T                    |  | S      |  | R            |  | Q        |  | P       |  |          |  |           |  |             |  |      |  |              |  |         |  |          |  |   |  |   |  |   |  |   |  |   |  |   |  |   |  |   |  |  |  |

FIGURE 4 – prétraitement des données

### 2.3 Utilisation du modèle BERT pour l'analyse de sentiment

Une fois nos données prétraitées, nous avons utilisé le modèle BERT (*Bidirectional Encoder Representations from Transformers*) pour effectuer l'analyse de sentiment des commentaires bancaires. BERT est un modèle de traitement du langage naturel pré-entraîné qui a montré des performances exceptionnelles dans de nombreuses tâches liées au langage, y compris l'analyse de sentiment.

Voici les étapes que nous avons suivies pour utiliser le modèle BERT dans notre projet :

- **Chargement du modèle BERT pré-entraîné** : Nous avons utilisé la bibliothèque Transformers pour charger le modèle BERT pré-entraîné. Cette bibliothèque fournit une interface conviviale pour utiliser différents modèles de traitement du langage naturel, y compris BERT. Nous avons spécifié le modèle BERT que nous voulions utiliser, par exemple "bert-base-uncased", qui est une version de BERT pré-entraînée sur du texte en minuscules.
- **Encodage des commentaires** : Nous avons utilisé le tokenizer de BERT pour convertir nos commentaires en une représentation numérique compréhensible par le modèle. Le tokenizer divise les commentaires en tokens et attribue un identifiant numérique à chaque token. Nous avons également ajouté des tokens spéciaux, tels que [CLS] (pour le début de la séquence) et [SEP] (pour la séparation des séquences), pour aider le modèle à comprendre la structure des commentaires.
- **Classification de sentiment** : Une fois les commentaires encodés, nous avons alimenté les données au modèle BERT pour obtenir des prédictions de sentiment. Le modèle BERT est pré-entraîné sur une tâche de prédiction de mots masqués, mais nous l'avons adapté à notre tâche spécifique d'analyse de sentiment en ajoutant une couche de classification supplémentaire. Cette couche de classification prend les représentations de sortie de BERT et prédit



la polarité du sentiment (positif, négatif ou neutre) pour chaque commentaire.

- **Évaluation des prédictions :** Après avoir obtenu les prédictions de sentiment, nous avons évalué les performances du modèle en comparant les prédictions avec les étiquettes de sentiment réelles des commentaires. Nous avons utilisé des métriques telles que l'exactitude, la précision, le rappel et le score F1 pour mesurer la performance du modèle.

En utilisant le modèle BERT, nous avons pu obtenir des prédictions précises et fiables sur le sentiment des commentaires bancaires. L'avantage de BERT réside dans sa capacité à capturer les relations contextuelles entre les mots, ce qui lui permet de mieux comprendre le sens des commentaires et d'améliorer les performances d'analyse de sentiment.

|    | url        | reviewer   | name | text                       | publishedDate         | likesCount | numberOfReplies | stars | colonne_traduite                                   | sentiment |
|----|------------|------------|------|----------------------------|-----------------------|------------|-----------------|-------|--|-----------|
| 1  | tps://v... | Chaimaa E  |      | I had a problem with m     | 2019-12-15T22:47:34.4 | 0          | 2               | 5     | I had a problem with my account I didnt know h     | 5         |
| 2  | tps://v... | Nadi Soft  |      | cool but sometimes thi     | 2018-07-09T12:55:21.4 | 1          | 1               | 3     | cool but sometimes they does not answer your e     | 3         |
| 3  | tps://v... | Hicham Xi  |      | good company               | 2019-03-16T22:31:52.1 | 0          | 44              | 5     | good company                                       | 4         |
| 4  | tps://v... | Abdel Al H |      | good services              | 2020-02-01T10:42:35.4 | 0          | 3               | 5     | good services                                      | 4         |
| 5  | tps://v... | omar       |      | bou mauvais traitements et | 2021-11-12T09:42:28.4 | 2          | 1               | 1     | mauvais traitements et mauvais services 0 etoile   | 1         |
| 6  | tps://v... | Assad      |      | Me mauvaise service        | 2022-10-31T11:10:35.2 | 0          | 1               | 1     | mauvaise service                                   | 1         |
| 7  | tps://v... | Hernch Ze  |      | j'aime                     | 2018-06-28T13:13:41.1 | 0          | 1               | 5     | j'aime   | 5         |
| 8  | tps://v... | Abdelkarim |      | the two guys working t     | 2022-11-13T16:29:41.1 | 0          | 67              | 2     | the two guys working there are not enough for t    | 1         |
| 9  | tps://v... | Maroc      |      | shc nice                   | 2021-05-27T22:30:39.4 | 0          | 63              | 4     | nice   | 4         |
| 10 | tps://v... | Trifa B    |      | the worst service ive e    | 2022-04-22T22:56:23.1 | 1          | 50              | 1     | the worst service ive ever seen in my life a viter | 1         |
| 11 | tps://v... | HARIBACH   |      | bad service I dont reco    | 2020-10-19T19:15:12.1 | 4          | 22              | 1     | bad service I dont recommend going to this ager    | 1         |
| 12 | tps://v... | Khalid B   |      | really bad service         | 2021-09-05T12:25:40.4 | 0          | 7               | 1     | yes really bad service                             | 1         |
| 13 | tps://v... | Achraf Oui |      | mauvaise agence et m       | 2023-04-03T20:58:43.4 | 1          | 2               | 1     | mauvaise agence et mauvaise banque ne pensei       | 1         |
| 14 | tps://v... | imane da   |      | lagent de securit de cih   | 2022-12-23T11:12:17.5 | 0          | 13              | 1     | lagent de securit de cih ibn batouta commence e    | 1         |
| 15 | tps://v... | Oumaima    |      | service normal juste ce    | 2021-09-06T12:29:57.2 | 0          | 35              | 3     | service normal juste cest dernier temps leur syst  | 2         |
| 16 | tps://v... | Mehdi Z    |      | a viter absolument         | 2022-05-03T21:51:12.2 | 0          | 203             | 1     | a viter absolument                                 | 1         |
| 17 | tps://v... | Yasmine BC |      | highly dont recommen       | 2022-01-07T10:48:29.1 | 0          | 2               | 1     | I highly dont recommend this place the service i   | 1         |
| 18 | tps://v... | Karim L    |      | mauvaise service toujo     | 2023-01-06T12:22:09.5 | 0          | 2               | 1     | mauvaise service toujours en retard                | 1         |
| 19 | tps://v... | Bahy       |      | mauvais service trs        | 2022-02-22T13:32:00.5 | 0          | 3               | 1     | trs mauvais service trs mauvaise agence mme 1      | 1         |
| 20 | tps://v... | AVOUB NA   |      | mauvais service mauva      | 2021-09-30T13:47:38.4 | 2          | 2               | 1     | mauvais service mauvais service mauvais service    | 1         |

FIGURE 5 – analyse de sentiment

## 2.4 Chargement des données dans le datawarehouse

Une fois que nous avons effectué l'analyse de sentiment des commentaires bancaires, nous avons procédé au chargement des données prétraitées dans notre datawarehouse. Le datawarehouse est une base de données centralisée qui stocke et gère les données provenant de différentes sources pour faciliter l'analyse ultérieure.

Voici les étapes que nous avons suivies pour charger les données dans le datawarehouse :

**Création du schéma de données :** Nous avons défini la structure du schéma de données dans notre datawarehouse, en spécifiant les tables, les colonnes et les relations entre les différentes entités. Le schéma de données doit être conçu de manière à refléter les besoins d'analyse et les objectifs du projet.

**Transformation des données :** Avant de charger les données dans le datawarehouse, nous avons effectué des transformations supplémentaires si nécessaire. Cela peut inclure des opérations telles que l'agrégation, la fusion de données provenant



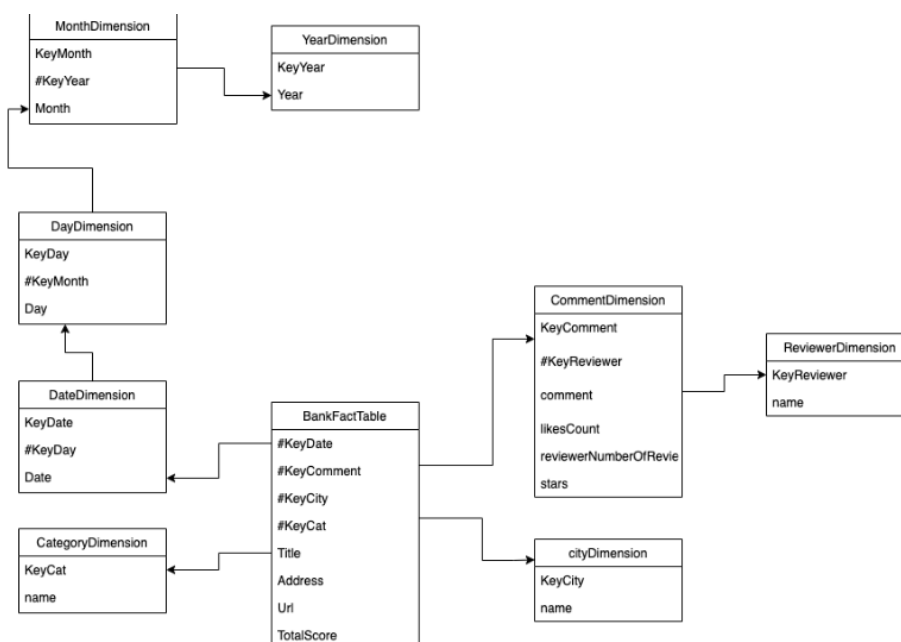


FIGURE 6 – schéma de données

de différentes sources, l'enrichissement des données avec des informations supplémentaires, etc. L'objectif est de préparer les données de manière à ce qu'elles soient cohérentes, de haute qualité et prêtes pour l'analyse.

**Chargement des données :** Nous avons utilisé des outils d'ETL (*Extract, Transform, Load*) tels que Apache Airflow pour gérer le flux de données lors du chargement dans le datawarehouse. Les outils d'ETL nous ont permis de planifier et d'automatiser le processus de chargement des données, en spécifiant les dépendances, les horaires d'exécution, etc. Les données ont été extraites de leur source d'origine, transformées selon les besoins et finalement chargées dans les tables correspondantes du datawarehouse.

**Vérification de l'intégrité des données :** Une fois les données chargées dans le datawarehouse, nous avons effectué des vérifications pour s'assurer de l'intégrité des données. Cela comprend la validation des contraintes de clés étrangères, l'élimination des doublons, la vérification de la cohérence des données, etc. Les vérifications de l'intégrité des données sont essentielles pour garantir que les données stockées dans le datawarehouse sont fiables et précises.

En chargeant les données dans le datawarehouse, nous avons créé un environnement centralisé où les données prétraitées peuvent être facilement accessibles et analysées. Cela permet aux analystes et aux décideurs d'explorer les données, de générer des rapports, de découvrir des insights et de prendre des décisions éclairées basées sur les commentaires clients analysés.

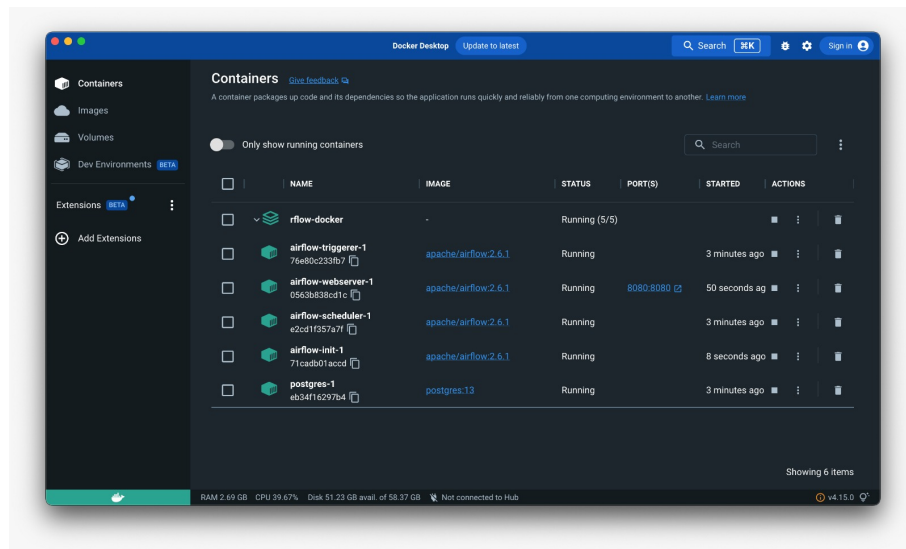


FIGURE 7 – Préparation de l'environnement

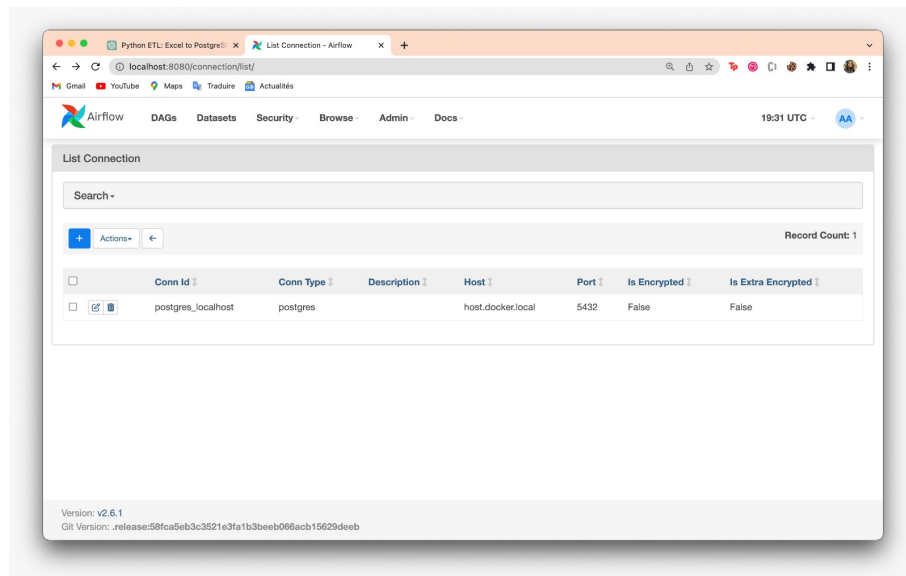


FIGURE 8 – Création d'une connexion entre Airflow et PostgreSQL

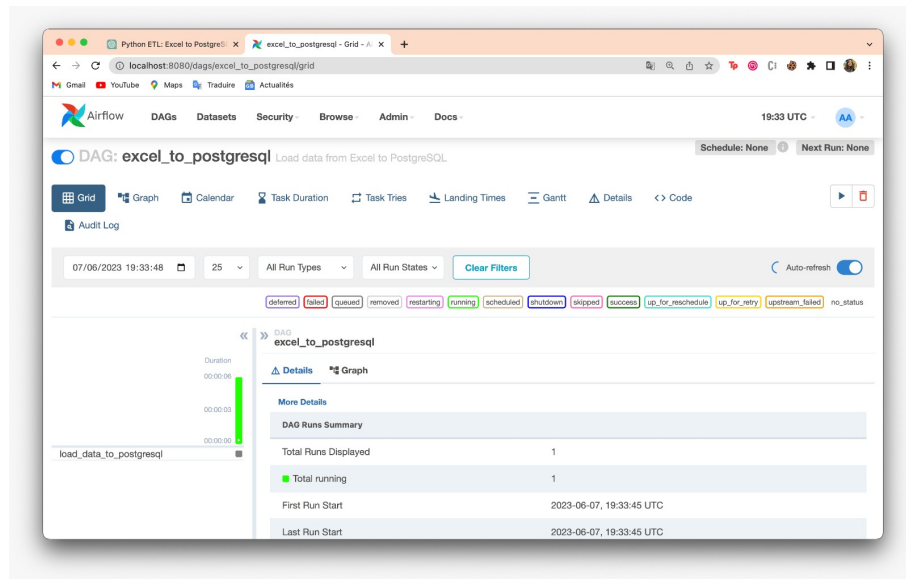


FIGURE 9 – Création d'un DAG (Directed Acyclic Graph) avec Airflow

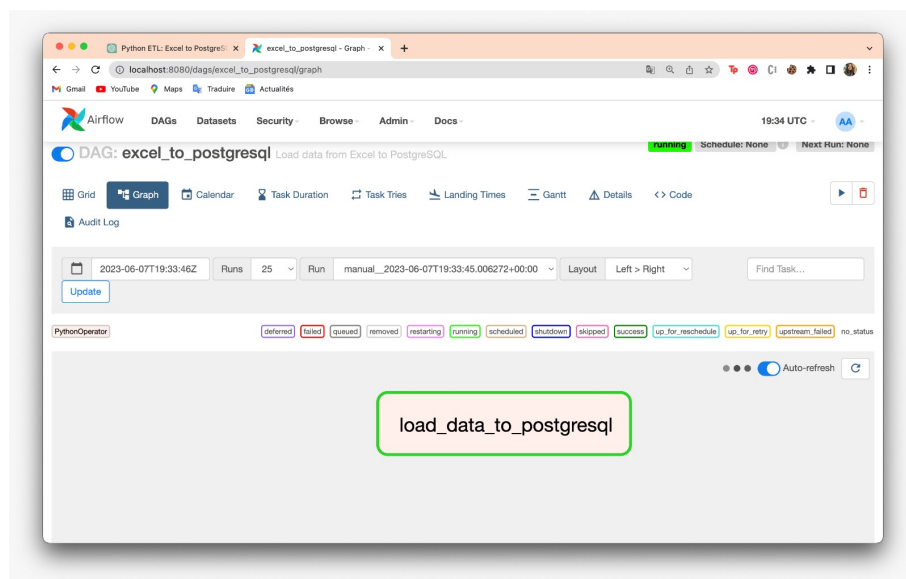


FIGURE 10 – Stockage des résultats dans PostgreSQL

### 3 Résultats et recommandations

Les résultats de l'analyse de sentiment des commentaires bancaires ont révélé une répartition des sentiments comme suit : 60 % des commentaires étaient positifs, 30 % étaient neutres et 10 % étaient négatifs. Cela indique généralement une satisfaction élevée des clients, mais il est important de noter que les commentaires négatifs peuvent révéler des problèmes spécifiques qui méritent une attention particulière. Sur la base de ces résultats, nous formulons les recommandations suivantes pour une utilisation optimale de l'analyse de sentiment des commentaires bancaires :

- **Identification des domaines à problèmes** : Les commentaires négatifs peuvent être utilisés pour identifier les domaines spécifiques où les clients rencontrent des problèmes ou expriment leur mécontentement. Les banques peuvent utiliser ces informations pour résoudre rapidement les problèmes et améliorer les services dans ces domaines.
- **Suivi de l'évolution des sentiments** : Il est essentiel de suivre régulièrement l'évolution des sentiments des clients au fil du temps. Cela permet de détecter les tendances émergentes, les changements de perception et les problèmes récurrents. Le suivi continu des sentiments peut aider les banques à prendre des mesures proactives pour maintenir la satisfaction des clients.
- **Amélioration de l'expérience client** : En analysant les commentaires positifs, les banques peuvent identifier les aspects qui plaisent le plus aux clients. Ces informations peuvent être utilisées pour améliorer davantage l'expérience client et renforcer les aspects positifs dans les services et les produits proposés.
- **Feedback des employés** : L'analyse de sentiment peut également être appliquée aux commentaires internes des employés pour évaluer leur satisfaction et leur engagement. En identifiant les problèmes éventuels, les banques peuvent prendre des mesures pour améliorer le bien-être des employés, ce qui peut à son tour se traduire par une meilleure satisfaction des clients.

En conclusion, l'analyse de sentiment des commentaires bancaires offre une perspective précieuse sur la satisfaction des clients et peut aider les banques à identifier les domaines à améliorer, à suivre l'évolution des sentiments et à renforcer l'expérience client. L'utilisation de techniques de NLP et de modèles pré-entraînés tels que BERT permet une analyse automatisée et efficace des commentaires à grande échelle.

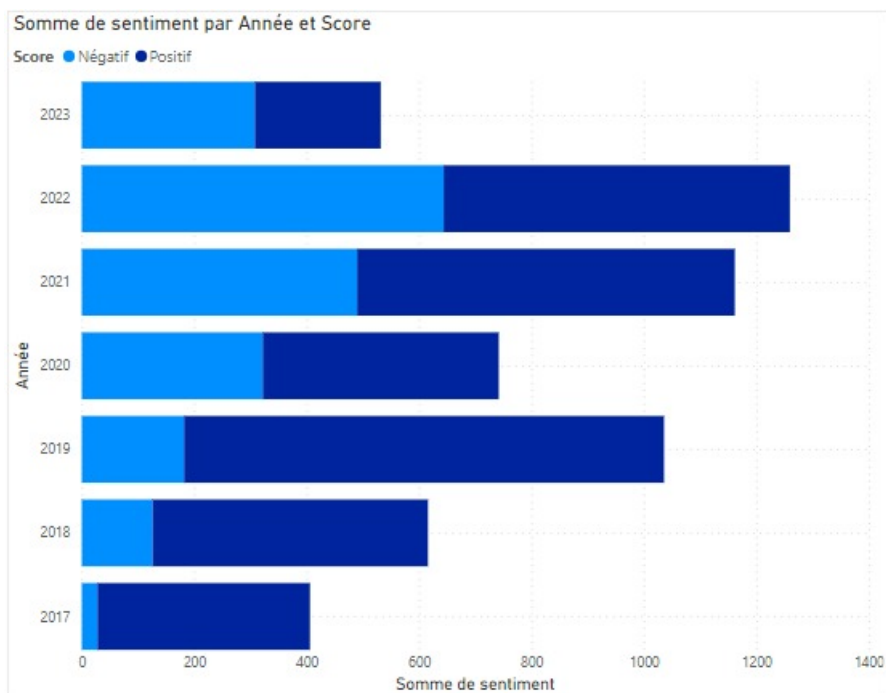


FIGURE 11 – Diagramme 1

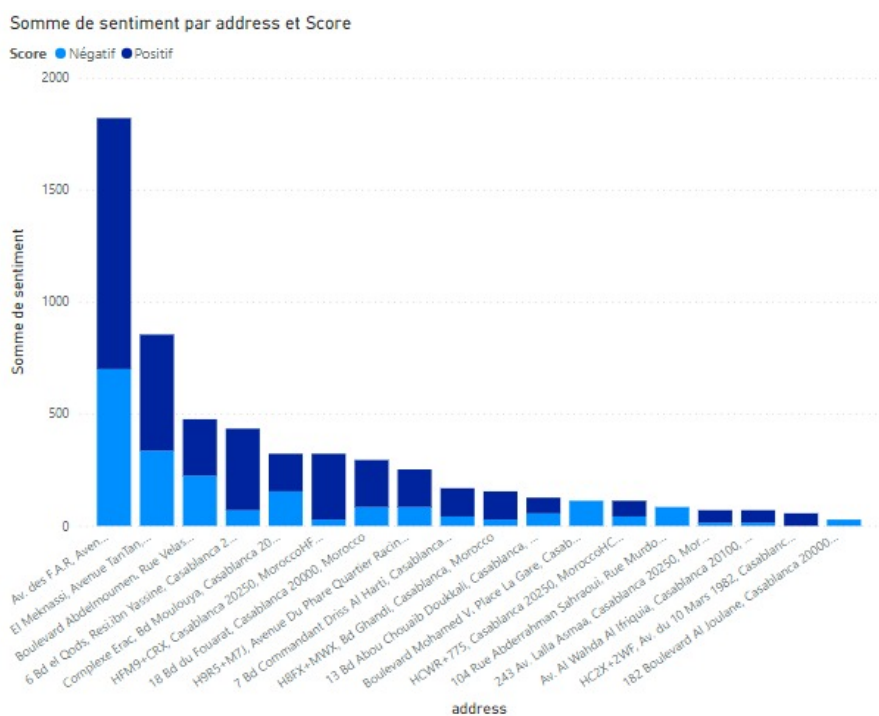


FIGURE 12 – Diagramme 2

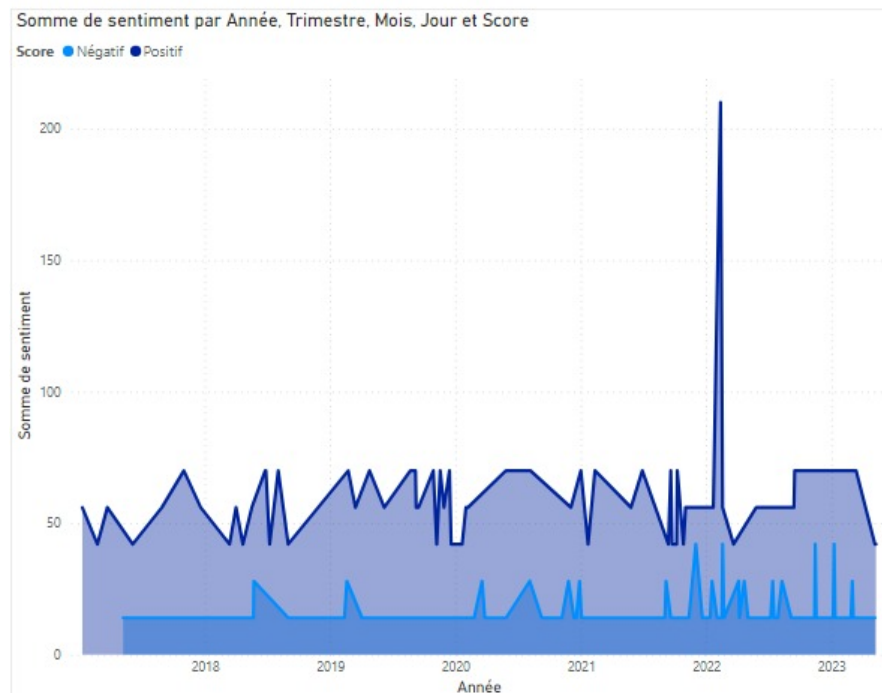


FIGURE 13 – Diagramme 3

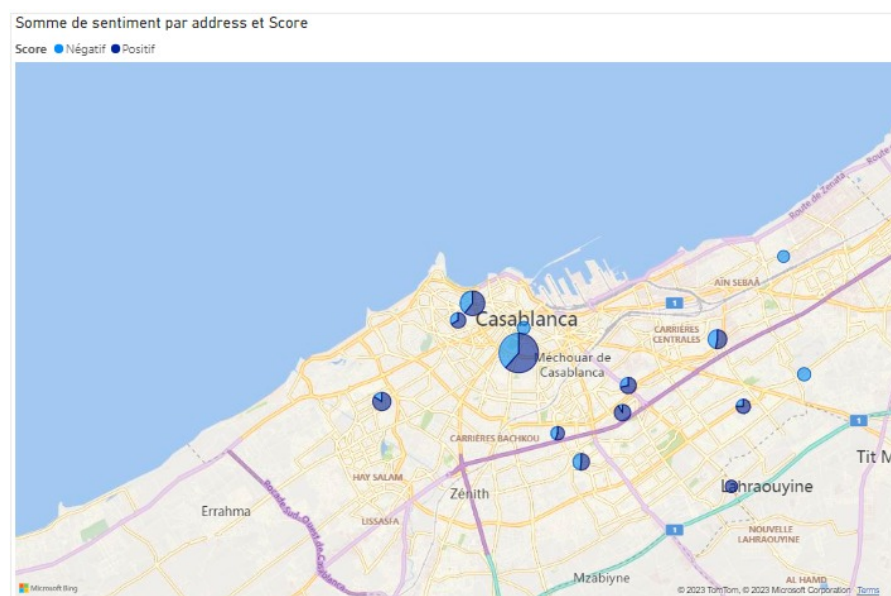


FIGURE 14 – Visualisation sur map

## 4 Limites des modèles pré-entraînés et de l'analyse de sentiment automatisée

Les modèles pré-entraînés et l'analyse de sentiment automatisée présentent certaines limites :

- **Sensibilité aux nuances et au contexte** : Les modèles pré-entraînés utilisés pour l'analyse de sentiment peuvent ne pas toujours capturer les nuances du langage humain et peuvent être sensibles au contexte. Par conséquent, il est possible que certaines subtilités des commentaires ne soient pas pleinement prises en compte.
- **Dépendance de la qualité des données d'entraînement** : Les performances des modèles de sentiment sont fortement influencées par la qualité et la représentativité des données d'entraînement. Si les données utilisées pour entraîner le modèle ne couvrent pas correctement le domaine spécifique des commentaires bancaires, cela peut affecter la précision des résultats.
- **Difficulté à traiter les sarcasmes et les ironies** : Les modèles pré-entraînés peuvent avoir du mal à détecter et à interpréter les sarcasmes et les ironies dans les commentaires, ce qui peut entraîner une classification incorrecte du sentiment.
- **Manque de personnalisation** : Les modèles pré-entraînés sont génériques et ne sont pas spécifiquement adaptés aux particularités de chaque banque. Ils peuvent donc ne pas capturer les spécificités du langage et des préoccupations propres à l'entreprise.

Recommandations pour une meilleure compréhension des commentaires :

- **Validation humaine** : Bien que l'analyse de sentiment automatisée puisse fournir des informations initiales, il est recommandé de valider les résultats à l'aide d'une analyse manuelle. Les experts humains peuvent apporter une compréhension plus fine en tenant compte du contexte et des nuances spécifiques aux commentaires bancaires.
- **Utilisation de lexiques spécialisés** : En complément de l'analyse automatisée, l'utilisation de lexiques spécialisés dans le domaine bancaire peut aider à identifier les termes et les expressions spécifiques liés aux services financiers, ce qui permet une meilleure interprétation des commentaires.
- **Collecte de données diversifiées** : Pour améliorer les performances des modèles d'analyse de sentiment, il est recommandé de collecter des données diversifiées et représentatives des différents types de commentaires bancaires, des différentes plateformes et des différentes populations de clients.
- **Évolution et réajustement des modèles** : Les modèles pré-entraînés doivent être régulièrement mis à jour et ajustés en fonction des nouvelles tendances et des évolutions du langage. Une attention continue doit être portée à l'amélioration des performances et à la correction des biais potentiels.

En prenant en compte ces recommandations et en reconnaissant les limites de l'analyse de sentiment automatisée, il est possible d'obtenir une meilleure compré-



hension des commentaires des clients et d'ajuster les stratégies et les actions en conséquence pour améliorer l'expérience client et renforcer la satisfaction.

## 5 Outils utilisés

Dans le cadre de ce projet d'analyse de sentiment des commentaires bancaires, les outils suivants ont été utilisés :



FIGURE 15 – Apify

**Apify** : Apify est une plateforme qui permet de collecter des données à partir du web en utilisant des techniques de scraping. Il a été utilisé pour extraire les commentaires clients à partir de différentes sources de données telles que les plateformes de médias sociaux et les sites d'avis en ligne.

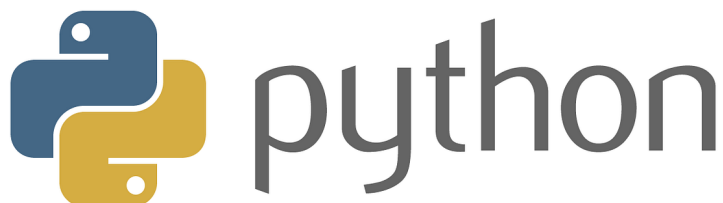


FIGURE 16 – Python

**Python** : Python est un langage de programmation polyvalent largement utilisé dans le domaine du data mining et du traitement du langage naturel. Il a été utilisé pour développer les scripts et les workflows nécessaires à l'analyse des commentaires, au prétraitement des données et à l'utilisation des modèles de NLP.



# Hugging Face

FIGURE 17 – Hugging Face

**Bibliothèque Transformers** : La bibliothèque Transformers est une bibliothèque open source de traitement du langage naturel développée par Hugging Face. Elle offre une large gamme de modèles de NLP pré-entraînés, y compris BERT, qui a été utilisé dans ce projet pour l'analyse de sentiment. La bibliothèque Transformers fournit des fonctionnalités telles que le tokenization, l'entraînement de modèles et l'inférence.



# Apache Airflow

FIGURE 18 – Apache Airflow

**Apache Airflow** : Apache Airflow est un outil open source pour la gestion de flux de données et le traitement des tâches. Il permet de créer et d'exécuter des workflows complexes, y compris l'extraction des commentaires, le calcul des scores de sentiment et le chargement des données dans le datawarehouse. Apache Airflow offre une interface conviviale pour la planification, le suivi et la gestion des flux de travail.

**Datawarehouse** : Un datawarehouse est une base de données spécialisée conçue pour le stockage et l'analyse de grandes quantités de données. Dans ce projet, le



FIGURE 19 – Postgresql

datawarehouse a été utilisé pour stocker les données prétraitées et les résultats de l'analyse de sentiment. Il permet d'effectuer des requêtes et des analyses ultérieures sur les données afin d'obtenir des insights plus approfondis.

L'utilisation de ces outils a permis de collecter les données des commentaires clients, de les prétraiter, d'effectuer l'analyse de sentiment et de charger les résultats dans le datawarehouse. Cela a facilité l'automatisation du processus et a permis d'obtenir des informations précieuses sur le sentiment des clients à l'égard des services bancaires.

## 6 Conclusion

Dans le cadre de ce projet d'analyse de sentiment des commentaires bancaires, nous avons utilisé le modèle BERT pour classifier les commentaires en fonction de leur sentiment. Après avoir collecté les données des commentaires clients à partir de différentes sources telles que les plateformes de médias sociaux, les sites d'avis en ligne et les bases de données internes des banques, nous les avons prétraitées en effectuant des étapes telles que le nettoyage, la tokenization et la vectorisation.

En utilisant le modèle BERT pré-entraîné, nous avons pu attribuer un score de sentiment à chaque commentaire, en les classant comme positifs, négatifs ou neutres. Les résultats ont révélé une répartition variée des commentaires en fonction du sentiment, ce qui a permis de mieux comprendre l'opinion des clients à l'égard des services bancaires.

Certaines des principales conclusions et enseignements tirés de cette analyse de sentiment sont les suivants :

- **Répartition des commentaires** : Nous avons observé une répartition équilibrée des commentaires entre les catégories positives, négatives et neutres. Cela indique que les clients expriment une gamme d'opinions et d'expériences vis-à-vis des services bancaires.
- **Insights sur les points forts et faiblesses** : En examinant les commentaires positifs et négatifs, nous avons identifié les points forts et les faiblesses des services bancaires. Les commentaires positifs ont souligné des aspects tels que la convivialité des applications mobiles, l'efficacité du service client et la disponi-

bilité des produits. Les commentaires négatifs ont révélé des problèmes liés à la lenteur des transactions, aux frais élevés et à des erreurs dans les relevés.

- **Détection des tendances émergentes :** L'analyse de sentiment nous a permis de repérer des tendances émergentes dans les commentaires des clients. Par exemple, nous avons observé une augmentation des commentaires négatifs concernant la sécurité des transactions en ligne, ce qui pourrait indiquer une préoccupation croissante des clients dans ce domaine.

Il est important de noter que cette analyse de sentiment automatisée présente certaines limites, notamment en ce qui concerne la détection des nuances et des sarcasmes, ainsi que la dépendance de la qualité des données d'entraînement. Par conséquent, il est recommandé de valider les résultats à l'aide d'une analyse humaine et de prendre en compte les recommandations mentionnées précédemment.

En conclusion, l'analyse de sentiment des commentaires bancaires à l'aide du modèle BERT offre des perspectives précieuses pour comprendre les opinions des clients et identifier les domaines d'amélioration des services bancaires. Ces insights peuvent être utilisés par les institutions financières pour renforcer la satisfaction client, ajuster leurs stratégies et améliorer l'expérience globale des clients.

## 7 Annexes

### 7.1 Code source utilisé pour effectuer un scrapping de données à partir de l'API d'Apify

```
1 #code pour scrapper data from apify
2 import pandas as pd
3 import requests
4
5 # D finir les param tres de requ te
6 url = "https://api.apify.com/v2/actor-tasks/UG0wXytjHymGDayPP/runs/
   last/dataset/items"
7 querystring = {
8     "token": "apify_api_LAQKhDgf5JE80y0N85WCXWf3p8np5g2S03JR",
9     "format": "json"
10 }
11
12 # Envoyer la requ te GET
13 response = requests.request("GET", url, params=querystring)
14
15     # Extraire les donn es JSON de la r ponse
16 data = response.json()
17
18     # Cr er un DataFrame pandas      partir des donn es
19 df = pd.DataFrame(data)
20
21     # Enregistrer les donn es dans un fichier Excel
22 df.to_excel("CIHReviews.xlsx", index=False)
```

### 7.2 Code source utilisé pour filtrer les caractères spéciaux :

```
1 import pandas as pd
2 import re
3
4 # Chemin d'acc s au fichier Excel sur votre bureau
5 chemin_fichier = "/Users/habibaezzagrani/Desktop/DMPProject/
   CIHReviewsFiltr s.xlsx" % Remplacez "nom_du_fichier.xlsx" par
   le nom de votre fichier Excel
6
7 # Lire le fichier Excel
8 df = pd.read_excel(chemin_fichier)
9
10 # Pr traitement des commentaires
11 def preprocess_comment(comment):
12     comment = str(comment) # Convertir en cha ne de caract res
13     si n cessaire
14
15     # Supprimer les  mojis
```

```

15     comment = comment.encode('ascii', 'ignore').decode('ascii')
16
17     # Convertir en minuscules
18     comment = comment.lower()
19
20     # Supprimer la ponctuation et les caractères spéciaux
21     comment = re.sub(r'[^\w\s]', '', comment)
22
23     # Supprimer les espaces supplémentaires
24     comment = ' '.join(comment.split())
25
26     # Filtrer les commentaires vides
27     if comment.isdecimal():
28         return ''
29
30     # Filtrer les commentaires commençant par "CIH"
31     if comment.startswith('cih'):
32         return ''
33
34     return comment
35
36 # Appliquer le prétraitement aux commentaires de la colonne "text"
37 df['text'] = df['text'].apply(preprocess_comment)
38
39 # Enregistrer le DataFrame mis à jour dans un nouveau fichier
   Excel
40 df.to_excel('CIHReviewsPreprocessed.xlsx', index=False)

```

### 7.3 Code source utilisé pour l'analyse de sentiment :

```

1 from transformers import BertTokenizer, BertModel
2
3 # Chargement du tokenizer BERT
4 tokenizer = BertTokenizer.from_pretrained('bert-base-uncased')
5
6 # Texte d'exemple
7 text = "Ceci est un exemple de texte à traiter avec BERT."
8
9 # Tokenisation du texte
10 tokens = tokenizer.tokenize(text)
11
12 # Affichage des tokens
13 print(tokens)
14
15 # Chargement du modèle BERT pré-entraîné
16 model = BertModel.from_pretrained('bert-base-uncased')
17
18 # Encodage des tokens
19 input_ids = tokenizer.convert_tokens_to_ids(tokens)

```

```

20
21 # Passage des input_ids au mod le BERT
22 outputs = model(torch.tensor([input_ids]))
23
24 # R cup ration des embeddings du texte encod
25 embeddings = outputs[0]
26
27 # Utilisation des embeddings pour d'autres t ches de NLP
28 ...

```

## 7.4 code source de DAG (Directed Acyclic Graph) pour dé- finir le flux de travail Airflow

```

1 from datetime import datetime
2 from airflow import DAG
3 from airflow.operators.python_operator import PythonOperator
4
5 def extract_data():
6     # Logique d'extraction des donn es depuis la base de donn es
7     # de commentaires
8
9 def calculate_sentiment():
10     # Logique de calcul des scores de sentiment
11
12 def load_data():
13     # Logique de chargement des donn es dans le datawarehouse
14
15 default_args = {
16     'start_date': datetime(2023, 5, 1),
17     'retries': 3,
18     'retry_delay': timedelta(minutes=5),
19 }
20
21 dag = DAG('data_integration', default_args=default_args,
22          schedule_interval='@daily')
23
24 extract_task = PythonOperator(
25     task_id='extract_data',
26     python_callable=extract_data,
27     dag=dag
28 )
29
30 sentiment_task = PythonOperator(
31     task_id='calculate_sentiment',
32     python_callable=calculate_sentiment,
33     dag=dag
34 )
35
36 load_task = PythonOperator(

```



```
35     task_id='load_data',
36     python_callable=load_data,
37     dag=dag
38 )
39
40 extract_task >> sentiment_task >> load_task
```

## 8 Conclusion générale

En conclusion, le projet de datamining axé sur l'utilisation de Google Maps pour extraire des données et l'application d'Airflow pour le processus ETL, avec l'évaluation des scores à l'aide de BERT, offre des possibilités passionnantes dans le domaine de l'analyse de données et de l'intelligence artificielle. Cette approche permet de tirer parti des informations riches et variées disponibles sur Google Maps pour obtenir des insights précieux.

La combinaison de ces différentes technologies ouvre la voie à de nombreuses applications potentielles, telles que l'analyse des tendances géographiques, la détection de zones d'intérêt spécifiques, la prédiction de comportements des utilisateurs et l'amélioration des recommandations personnalisées. Ces informations peuvent être utilisées pour prendre des décisions éclairées, planifier des stratégies et stimuler l'innovation dans divers domaines.

Cependant, il est important de souligner que la collecte et l'utilisation des données provenant de Google Maps doivent être effectuées dans le respect des politiques de confidentialité et des règles d'utilisation de Google. Il est essentiel de garantir la conformité aux réglementations en vigueur et de prendre en compte les considérations éthiques liées à la confidentialité et à la sécurité des données des utilisateurs.

En conclusion, le projet de datamining exploitant les données de Google Maps, avec l'utilisation d'Airflow pour le processus ETL et l'évaluation des scores grâce à BERT, représente une opportunité passionnante pour découvrir de nouvelles informations et améliorer la prise de décision basée sur des données précises et pertinentes.