

Rapport de projet : Base de données

Cosmétovigilance

Master 2 EID²

30 Janvier 2025

Présenté par : Amadou O. Diallo

Aminata A. Ba

Yakaré Koité

Résumé

Ce projet vise à identifier et analyser les risques potentiels liés à l'utilisation des produits cosmétiques en détectant les ingrédients incompatibles ou nocifs. Grâce à une base de données relationnelle, nous avons structuré les informations sur les produits, ingrédients, et leurs interactions. Des requêtes SQL ont permis de repérer les produits contenant des composants dangereux ou similaires afin d'assurer la sécurité des consommateurs.

L'objectif final est d'offrir un outil performant pour l'analyse et la prévention des risques dans l'industrie cosmétique.

Table des matières

1	Introduction	-2
2	La cosmétovigilance	-2
3	Le scraping ou la collecte des données	-1
4	Les métadonnées	-1
5	La modélisation de la base de données	1
5.1	Le modèle entité-association	1
5.2	Le modèle relationnel	2
6	Le traitement des données	3
6.1	Nettoyage des données	3
6.2	La normalisation des données : Troisième Forme Normale	4
6.3	Gestion des données similaires :	4
6.3.1	Traitement des ingrédients similaires :	4
6.3.2	Analyse des compositions des produits	4
6.4	Résolution des problèmes de relations entre les tables	5
6.5	Mise en place de scripts Python	5
6.5.1	ETL (Extraction, Transformation, Chargement) :	5
6.6	Déclaration des ingrédients nocifs	5
7	Le site web	6
7.1	Le front-end	7
7.1.1	Technologie Utilisée	7
7.1.2	Pages Créées	7
7.2	Le Back-end	8
7.2.1	Technologie Utilisée	8
7.2.2	Fonctionnalités Back-End Implémentées	8
8	L'état d'avancement du projet actuel (stade actuel)	8
9	Les difficultés rencontrées	9
9.1	La difficulté d'accès aux données :	9
9.2	Les défis du prétraitement des données :	9
9.3	L'intégration des données :	9
9.4	Identification des interactions néfastes pour les ingrédients	10

10 Les solutions mises en place	10
10.1 Traitement des listes d'ingrédients :	10
10.2 Gestion des doublons	11
10.3 Création de la table des produits similaires	11
10.4 Remplissage de la table interactions	11
10.5 Statistique	12
11 Perspectives	12
12 Conclusion	12
13 Références	13

1 Introduction

La cosmétovigilance est une discipline essentielle qui consiste à surveiller les effets indésirables liés à l'utilisation de produits cosmétiques. Avec l'augmentation constante du nombre de produits sur le marché, la collecte et l'analyse des données associées à leurs ingrédients deviennent cruciales pour garantir la sécurité des consommateurs. Dans ce projet, nous utilisons des techniques de scraping et d'extraction de données pour enrichir une base de données centralisée, permettant de faciliter la recherche d'informations sur les produits et ingrédients cosmétiques. L'objectif principal est de modéliser, traiter et organiser ces données tout en construisant un site web accessible au public, combinant des fonctionnalités de recherche avancées et une interface utilisateur intuitive. Nous abordons également les défis liés à la gestion des doublons, des données similaires et des analyses des interactions potentiellement néfastes entre ingrédients.

Ce rapport vise à présenter un aperçu complet de toutes les tâches qui ont été effectuées lors de ce projet. Il est organisé comme suit : La section 2 donne une explication complète de la cosmétovigilance. La section 3 présente la manière dont nous avons collecté les données sur les sites de produits de beauté. La section 4 présente les métadonnées de notre BDD. La section 5 présente la structure et les différents éléments qui ont servi à modéliser notre base de données. La section 7 présente l'interface de notre site web. La section 8 présente l'avancement du projet et l'état actuel. Et enfin les sections 9, 10 présentent les difficultés rencontrées ainsi que les solutions adoptées lors de la réalisation de ce projet.

2 La cosmétovigilance

Lors de ce projet notre but était de surveiller et prévenir les risques liés à l'utilisation des produits cosmétiques. Cela inclut :

- La détection d'effets secondaires inattendus.
- L'évaluation des interactions entre ingrédients.
- La mise en place de bases de données pour le signalement des problèmes liés aux cosmétiques.

Après une base de données mise en place, nous avons inspecté les ingrédients afin de voir celles qui sont nocives entre elles. Un produit cosmétique est constitué de plusieurs ingrédients actifs et complémentaires, chacun ayant un rôle bien précis : hydrater la peau, améliorer la texture, protéger ou conserver la formule. Toutefois, certaines associations d'ingrédients peuvent être déconseillées en raison de leur impact sur la peau ou de leur incompatibilité chimique. C'est donc suivant cette logique que nous avons travaillé sur ce projet.

3 Le scraping ou la collecte des données

Le scraping est une méthode d'extraction de données automatisée depuis des sites web. Dans ce projet, nous avons identifié plusieurs plateformes clés contenant des informations pertinentes sur les produits cosmétiques et leurs ingrédients. Ces données comprennent les descriptions, les listes d'ingrédients, les avis utilisateurs et les éventuels effets indésirables. Nous avons ciblé des plateformes spécialisées dans la cosmétique, telles que :

— LOOKFANTASTIC

— COSMETIFY

— SEPHORA

— BEAUTY BAY

4 Les métadonnées

Les métadonnées jouent un rôle central dans l'organisation des informations collectées. Elles permettent de décrire les données scrappées avec des attributs comme : Le nom du produit. La catégorie (crème, lotion, maquillage, etc.). Les ingrédients principaux et leurs concentrations. Les avis des utilisateurs ou des experts. Nous vous présentons notre métadonnée :

Table	Colonnes	Clés	Relations	Contraintes/Autres
PRODUCTS	<ul style="list-style-type: none"> — ID_P — P_NAME — P_SIZE — P_PRICE — P_BRAND — DESCRIPTION — DEVISE (€, \$, etc.) — SIZE_TYPE (ml, g) 	Clé primaire : ID_P	<ul style="list-style-type: none"> — Liée à SIMILAR_PROD via ID_P — Liée à COMPOSITION via ID_P 	Contraintes : <ul style="list-style-type: none"> — DEVISE (€, \$, etc.) — SIZE_TYPE (ml, g)
INGREDIENTS	<ul style="list-style-type: none"> — ID_I — I_NAME 	Clé primaire : ID_I	<ul style="list-style-type: none"> — Liée à COMPOSITION via ID_I — Liée à INTERACTIONS via ID_I1 et ID_I2 — Liée à SIMILAR_ING via ID_I 	-
COMPOSITION	<ul style="list-style-type: none"> — ID_P (prod) — ID_I (ingr) 	Clé primaire : (ID_P, ID_I)	<ul style="list-style-type: none"> — FK vers PRODUCTS(ID_P) — FK vers INGREDIENTS(ID_I) 	Relation produit-ingrédient
INTERACTIONS	<ul style="list-style-type: none"> — ID_I1 — ID_I2 — HARMFUL (0 ou 1) — DESCRIPTION 	Clé primaire : (ID_I1, ID_I2)	<ul style="list-style-type: none"> — FK vers INGREDIENTS(ID_I) 	Contrainte : <ul style="list-style-type: none"> — HARMFUL (0 ou 1)
SIMILAR_PROD	<ul style="list-style-type: none"> — ID_SP — ID_P — SP_NAME 	Clé primaire : ID_SP	<ul style="list-style-type: none"> — FK vers PRODUCTS(ID_P) 	Identifie des produits similaires
SIMILAR_ING	<ul style="list-style-type: none"> — ID_SI :PK — ID_I :FK_ing — SI_NAME 	Clé primaire : ID_SI	<ul style="list-style-type: none"> — FK vers INGREDIENTS(ID_I) 	Identifie des ingrédients similaires
FEEDBACK	<ul style="list-style-type: none"> — ID_F :PK — F_NAME — F_BRAND — AVIS 	Clé primaire : ID_F	—	Permet aux utilisateurs de donner leur avis

TABLE 1 – Description des tables et relations dans la base de données.

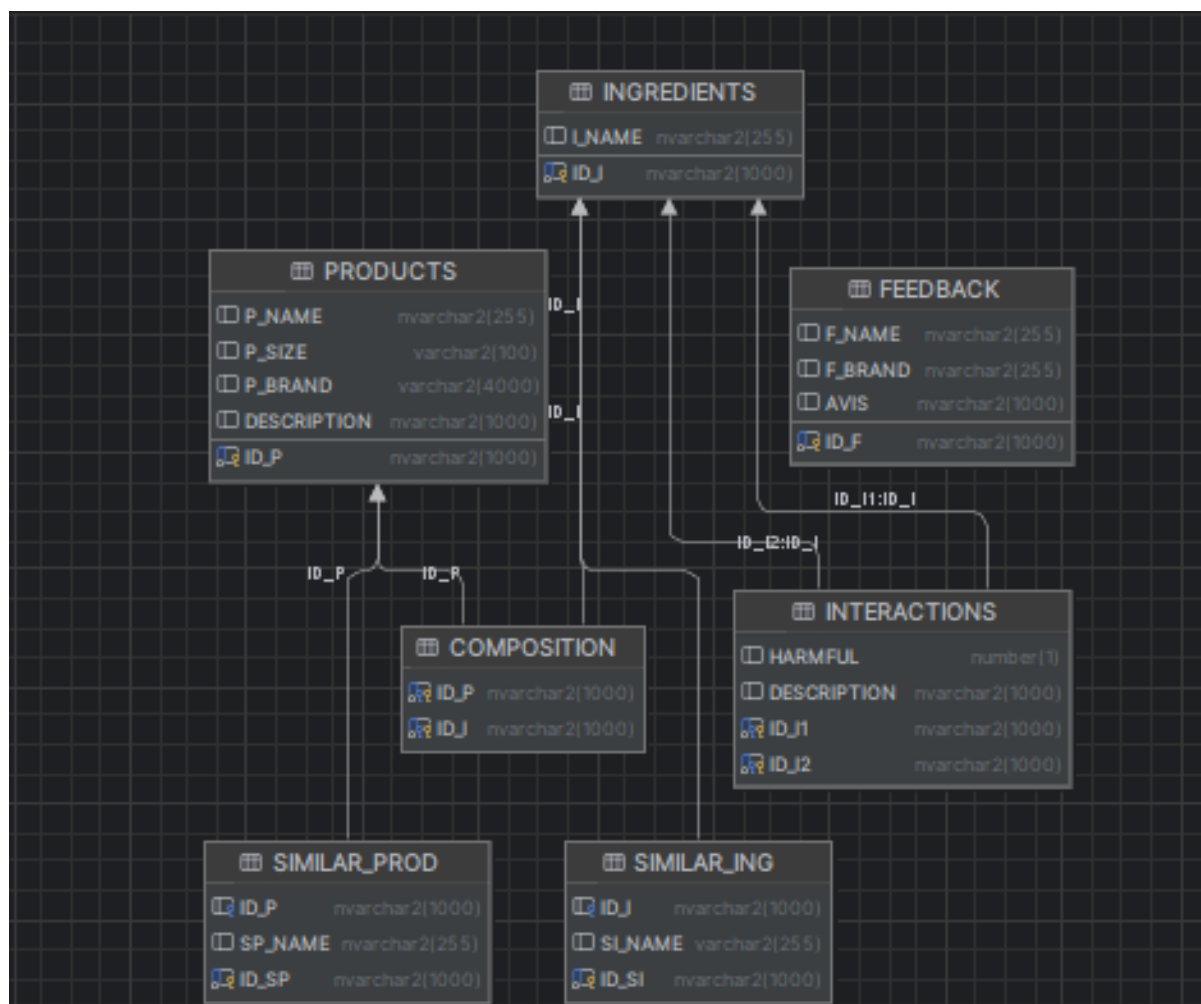
5 La modélisation de la base de données

Le MEA permet de représenter les entités principales, PRODUCTS et INGREDIENTS, ainsi que leurs relations, telles que la composition et les interactions entre ingrédients. Cette modélisation graphique met en évidence les dépendances et les cardinalités pour structurer les données de manière logique. Le passage au MRD traduit ces concepts en une structure relationnelle, où chaque entité et relation devient une table dans la base de données, avec des clés primaires (comme ID_P et ID_I) et des contraintes assurant l'intégrité des données. Ce processus garantit une base de données cohérente et fonctionnelle, prête pour des applications pratiques.

5.1 Le modèle entité-association

Le schéma conceptuel de base de données illustre une relation entre des produits et leurs ingrédients, avec des liens décrivant leur composition et leurs interactions. Chaque produit est composé d'un ou plusieurs ingrédients, et des interactions spécifiques entre ces ingrédients peuvent être décrites, notamment en cas d'effets nuisibles. Le modèle inclut également des relations de similarité, permettant de comparer des produits ou des ingrédients ayant des caractéristiques communes. Les cardinalités précisent les contraintes quantitatives sur ces relations, assurant une structure cohérente pour la gestion des données.

5.2 Le modèle relationnel



La relation entre les produits et leurs ingrédients est gérée par la table COMPOSITION, qui utilise les clés étrangères ID_P et ID_I pour associer chaque produit à ses ingrédients. Une table spécifique, INTERACTIONS, est utilisée pour définir les interactions possibles entre deux ingrédients, identifiés par ID_I1 et ID_I2. Elle inclut des informations sur le caractère nocif de l'interaction et une description de celle-ci. Enfin, les tables SIMILAR_PROD et SIMILAR_ING servent à gérer les similarités respectivement entre les produits et entre les ingrédients. Elles utilisent des clés spécifiques (ID_SP pour les produits similaires et ID_SI pour les ingrédients similaires) et associent des noms et descriptions correspondants. L'ensemble de ces tables et relations permet une gestion complète des produits, de leurs compositions, des interactions chimiques, et des produits ou ingrédients similaires.

6 Le traitement des données

Pour garantir une base de données cohérente, fiable et exploitable, plusieurs étapes ont été mises en place pour traiter et organiser les données collectées. Ces étapes incluent des processus de nettoyage, de normalisation et d'organisation.

Colonne	Problème	Nombre de lignes	Commentaire
Produit – Avant prétraitement			
P_Name	Valeurs nulles <code>name</code>	20	Nom manquant
INGREDIENTS	Valeurs nulles <code>ingredients</code>	67	Ingrédients manquants
DESCRIPTION	Valeurs nulles dans <code>description</code>	4001	Description manquante
P_PRICE	Valeurs nulles dans <code>prix</code>	39	Prix non défini
P_SIZE	Valeurs nulles dans <code>taille</code>	143	Taille absente
*	Doublons	1990	Lignes dupli- quées
*	Lignes après suppression des null	9426	Lignes valides après traitement
*	Lignes après suppression des doublons	7436	Total final

TABLE 2 – Récapitulatif du nettoyage de la table Produit

6.1 Nettoyage des données

1. **Suppression des doublons** : Identification et suppression des enregistrements redondants dans les tables des produits et des ingrédients. Utilisation de scripts SQL pour comparer les enregistrements basés sur des clés uniques.
2. **Correction des erreurs de format** : Harmonisation des noms d'ingrédients (casse uniforme, suppression des caractères spéciaux). Vérification et mise à jour des identifiants incorrects.

6.2 La normalisation des données : Troisième Forme Normale

1. **Standardisation des formats** : Unification des formats de prix (conversion en euros ou dollars selon un taux prédéfini). Transformation des listes d'ingrédients séparées par des virgules, des espaces ou d'autres délimiteurs en une structure standard.
2. **Création de colonnes spécifiques** : Ajout de colonnes dédiées pour indiquer l'unité des prix et gérer les données hétérogènes.
3. **Après la création des tables ingrédient, composition, la colonne ingrédients de produit est supprimer pour respecter les propriété normale.**
4. Nous avons choisi de laisser les colonne size et price contenir des valeur null car sans cela le dataset diminuera en nombre de lignes...

6.3 Gestion des données similaires :

6.3.1 Traitement des ingrédients similaires :

1. Développement de scripts pour identifier les similitudes entre les noms d'ingrédients en utilisant des algorithmes comme edit_distance, jaro_winkler, soundex entre autre.
2. Création d'une table dédiée aux ingrédients similaires basée sur un seuil de similarité défini.

Voici un exemple : l'engredient AQUA rien que lui compte 9 similaires

	similars	ED	JW	s1	s2
AQUA	AQUA (PURIFIED WATER)	17	0.838095238095238	A200	A216
AQUA	AQUA (WATER)	8	0.819512195121951	A200	A236
AQUA	AQUA (WATER/EAU)	12	0.85	A200	A236
AQUA	AQUA/ WATER/ EAU	12	0.85	A200	A236
AQUA	WATER\AQUA\EAU	14	0.424074074074074	A200	G526
AQUA	AQUA [WATER]	8	0.494444444444444	A200	A235
AQUA	AQUA[WATER]	7	0.426767676767676	A200	A235
AQUA	PURIFIED WATER / AQUA	19	0.531746031746031	A200	P613
AQUA	WATER (AQUA/EAU)	12	0.666666666666666	A200	W362

6.3.2 Analyse des compositions des produits

Vérification des interactions entre ingrédients potentiellement problématiques.

6.4 Résolution des problèmes de relations entre les tables

1. **Élimination des cycles dans les relations** : Restructuration des tables pour éviter les dépendances circulaires.
2. **Optimisation des relations** : Définition de clés primaires et étrangères pour garantir une intégrité référentielle solide.

6.5 Mise en place de scripts Python

Nous avons eu à implémenter des script python pour le scraping mais aussi pour automatisé certaines taches comme la génération des requêtes de mise à jour (UPDATE) de la table interaction et les requête d'insertion des tables similar_prod et similar_ing.

6.5.1 ETL (Extraction, Transformation, Chargement) :

1. Nous avons extrait les données des sites, puis les avons transformé en ligne de fichier CSV et pui les avons chargé dans la base de données
2. Scripts Python développés pour extraire les données des sites scrappés, transformer leur structure et les charger dans la base de données.
3. Automatisation des tâches de nettoyage et de normalisation des données.

6.6 Déclaration des ingrédients nocifs

Il a fallu identifier les ingrédients qui sont nocifs entres elles et déclaré cela dans la base de données.

7 Le site web


Cosmétovigilance

Votre plateforme pour analyser les ingrédients, produits et avis.

Bienvenue

Cosmétovigilance vous aide à découvrir et analyser les composants des produits cosmétiques.
Explorez notre plateforme pour mieux comprendre les ingrédients, évaluer les produits et partager vos avis.


Analysez les Ingrédients



Identifiez les composants des produits cosmétiques et leur impact potentiel.

Explorer les Ingrédients


Découvrez les Produits



Recherchez des produits spécifiques et consultez leurs détails complets.

Explorer les Produits

Partagez vos Avis



Exprimez votre opinion sur vos produits préférés ou partagez des retours d'expérience.

Donner un Avis

© 2024 Cosmétovigilance B08. Tous droits réservés.

Cosmétovigilance

Recherche d'ingrédients et de produits

Nom de l'ingrédient :

Aucun ingrédient trouvé. Vous cherchez peut-être?

AQUA

Rechercher

Nom du produit :

SENSUOUS ESSENTIAL OIL BLEND

SLEEP EASY PILLOW AND ROOM MIST

SPRAY THERMOPROTECTEUR

SPRAY THERMOPROTECTEUR

Rechercher

© 2024 Cosmétovigilance B08

← Retour à l'accueil

Analyse des Ingrédients

Recherchez des ingrédients et découvrez leurs impacts sur vos produits cosmétiques.

Nom de l'ingrédient :

Rechercher

Résultats de la recherche

ANHIDROXYLITOL

XYLITOL

Produits contenant cet ingrédient :

ALL BODY BUTTER

ATODERM BODY WASH

HIGH-TEA MOISTURE MULTIPURPOSE BALM

NATURAL MOISTURIZING FACTORS + INULIN BODY LOTION

NOURISHING CURL ENHANCER

DEDECTIM/3 DOIMED HEAT PROTECTION SODAY

Retour à l'accueil

Découverte des Produits

Explorez notre base de données pour trouver des produits et leurs caractéristiques.

Nom du produit :

MINI SCENTED CANDLE TRIO

Rechercher

Résultats de la recherche

MINI SCENTED CANDLE TRIO

Ingrédients :

ALOE VERA

ARACHIDIC ACID

BOSWELLIA CARTERII OIL

FRANKINCENSE & SANDALWOOD CANDLE

GLYCERYL STEARATE

HYDROGENATED SOYBEAN OIL

Retour à l'accueil

Partagez vos Avis

Donnez votre avis

Nom du produit :

Taper un produit

Marque :

Marque du produit

Avis :

Votre avis...

Soumettre

EAU MICELLAIRE CHANEL: CHANEL

J'ai voulu tester ce produit pour me maquiller mais j'ai eu une reaction allergique.

Publié récemment

MINI SCENTED CANDLE TRIO: MAYBELLINE

Produit ayant des effets nefastes sur la peau.

Publié récemment

ELIXIR VOLUME AND GROWTH SPRAY: PANTENE

Ce produit a des effets nefastes.

Publié récemment

BB CREAM NIVEA PREMIUM: NIVEA

Ce produit provoque des boutons sur la peau. Il contient de l'acide lactique et j'ai eu une reaction allergique.

Publié récemment

7.1 Le front-end

7.1.1 Technologie Utilisée

1. **HTML/CSS** : Structuration et stylisation des pages, avec un design simple et fonctionnel.
2. **JavaScript et Ajax** : Gestion des interactions dynamiques sur les pages, comme les recherches en temps réel ou les formulaires interactifs, la recuperation du contenu des champs de saisie entre autre.

7.1.2 Pages Créées

1. **Recherche de produits** : Permettre aux utilisateurs de rechercher des informations sur des produits cosmétiques. On a fait une barre de recherche avec auto-complétion. Ensuite on affiche les détails du produit : Ingrédients, interactions connues, description. On suggère également des produits similaires en cas de recherche infructueuse.

2. **Recherche des ingrédients** : Permettre aux utilisateurs de rechercher des informations sur des ingrédients. On a fait une barre de recherche avec auto-complétion. Ensuite on informe les utilisateurs sur les interactions potentielles entre les ingrédients.
3. **Avis** : Permettre aux utilisateurs de donner leur avis sur les produits qu'ils ont utilisé.

7.2 Le Back-end

7.2.1 Technologie Utilisée

1. **Langage PHP** : Utilisé pour la gestion de la logique serveur et l'interfaçage (interroger le serveur) avec la base de données.
2. **Base de Données : Oracle Database** Stockage des produits, des ingrédients, des interactions, et des déclarations d'effets indésirables.
3. **Architecture** : Une approche structurée avec des scripts PHP dédiés pour différentes fonctionnalités (gestion des produits, des ingrédients, et des interactions).

7.2.2 Fonctionnalités Back-End Implémentées

1. **Gestion des Produits** : Requête SQL pour récupérer les détails d'un produit selon son nom ou ses caractéristiques. Suggestions automatiques en cas de recherche infructueuse. Recherche basée sur la similarité phonétique (SOUNDEX) et les distance d'édition (edit_distance, Jaro_winkler). Gestion des interactions entre les produits et les ingrédients. Affichage de la liste des produits contenant des ingrédients incompatibles. Identification des interactions néfastes.
2. **Gestion des Ingrédients** : Requêtes pour récupérer tous les ingrédients associés à un produit. Recherche des interactions entre ingrédients. Affichage des descriptions des interactions (description des effets négatifs potentiels).

8 L'état d'avancement du projet actuel (stade actuel)

Actuellement, le projet est à un stade avancé. La structure des bases de données a été finalisée, incluant la table des produits, des ingrédients, des compositions, des ingrédients similaires, des produits similaires et des interactions entre ingrédients. En complément, un site statique de recherche a été développé, permettant aux utilisateurs de rechercher

des ingrédients et des produits, offrant une base fonctionnelle pour explorer les données et leurs interconnexions.

9 Les difficultés rencontrées

Lors de la réalisation de ce projet, plusieurs défis majeurs ont été rencontrés, notamment dans les étapes de collecte, de prétraitement et d'intégration des données. Voici les principaux obstacles identifiés :

9.1 La difficulté d'accès aux données :

L'un des premiers défis a été de trouver des données pertinentes sur les produits cosmétiques. Bien que de nombreux sites web offrent des informations utiles, beaucoup d'entre eux imposent des restrictions, comme un accès payant ou des limitations de scraping. Cela a nécessité de faire des choix parmi les plateformes disponibles et d'adopter des stratégies spécifiques pour accéder à un volume suffisant de données.

9.2 Les défis du prétraitement des données :

Le prétraitement des données a également été une étape complexe, notamment lors de la création de la table des ingrédients. Les principales difficultés incluent :

1. Choix du seuil de similarité des ingrédients
2. La variation des séparateurs : Certains ingrédients étaient listés séparément par des virgules, tandis que d'autres étaient séparés par des points-virgules, des espaces, des points. Cette hétérogénéité a rendu difficile la création d'une table uniforme et a nécessité des traitements spécifiques pour harmoniser les formats.
3. Les incohérences dans la structure des données : Parfois, des informations importantes sur les ingrédients étaient regroupées ou manquantes, rendant l'organisation des données plus laborieuse.

9.3 L'intégration des données :

L'intégration des données collectées dans la base de données centralisée a également posé plusieurs problèmes : Les données étant scrappées sur des plateformes différentes, leur structure était différente d'une plateforme à l'autre, ce qui est très difficile à prétraiter et à combiner dans une seule base de données. Les données manquantes : Certaines informations, comme les prix, les tailles ou les concentrations des ingrédients, étaient incomplètes, ce qui a nécessité des ajustements pour les traiter sans compromettre la qualité globale de la base. L'hétérogénéité des formats : Les prix, par exemple, étaient exprimés

dans différents formats : certains en euros, d'autres en dollars américains (USD) et parfois dans d'autres devises non identifiées. Cela a nécessité un processus de conversion et de standardisation pour uniformiser les données.

9.4 Identification des interactions néfastes pour les ingrédients

L'un des principaux défis rencontrés était la prise de décision concernant la toxicité ou l'interaction néfaste entre deux ingrédients spécifiques. Il fallait développer une méthodologie ou utiliser des sources fiables pour déterminer si une paire d'ingrédients avait une interaction potentiellement néfaste. Cela nécessitait une analyse approfondie basée sur : des bases de données scientifiques existantes sur les interactions chimiques ou biologiques entre ingrédients et des études toxicologiques ou dermatologiques liées à l'usage combiné de certains ingrédients dans des formulations cosmétiques ou pharmaceutiques. Avec un total de 35 000 combinaisons possibles d'ingrédients, remplir une table aussi volumineuse manuellement aurait été inefficace et fastidieux.

Malgré ces difficultés, des solutions ont été mises en place pour surmonter ces obstacles, notamment en utilisant des techniques avancées de traitement des données et des scripts spécifiques pour nettoyer et harmoniser les informations collectées.

10 Les solutions mises en place

Pour surmonter les différentes difficultés rencontrées au cours du projet, plusieurs solutions ont été mises en place, permettant de répondre efficacement aux problèmes de collecte, de prétraitement et d'intégration des données. Voici les principales stratégies adoptées :

10.1 Traitement des listes d'ingrédients :

Face à l'hétérogénéité des séparateurs utilisés dans les listes d'ingrédients (virgules, point-virgule, espaces, etc.), nous avons développés des procédures PL/SQL spécifiques pour traiter chaque groupe de cas de manière adaptée. Ces scripts ont permis :

1. Supprimer certaines données scrapés comme les notes (avis) d'un produit
2. De détecter automatiquement les séparateurs utilisés dans les données.
3. De standardiser les listes d'ingrédients en les scindant de manière cohérente.
4. D'enrichir la table des ingrédients tout en évitant les erreurs ou les omissions.
5. Choix de plusieurs seuil, tester puis valider

10.2 Gestion des doublons

L'élimination des doublons a été une étape clé pour garantir la qualité et la cohérence de la base de données. Pour cela, nous avons créés block PL/SQL de déduplication pour identifier les entrées répétées dans les tables des produits et des ingrédients. Nous avons éliminé les doublons en nous basant sur des clés uniques, comme le nom du produit, la catégorie ou les métadonnées associées.

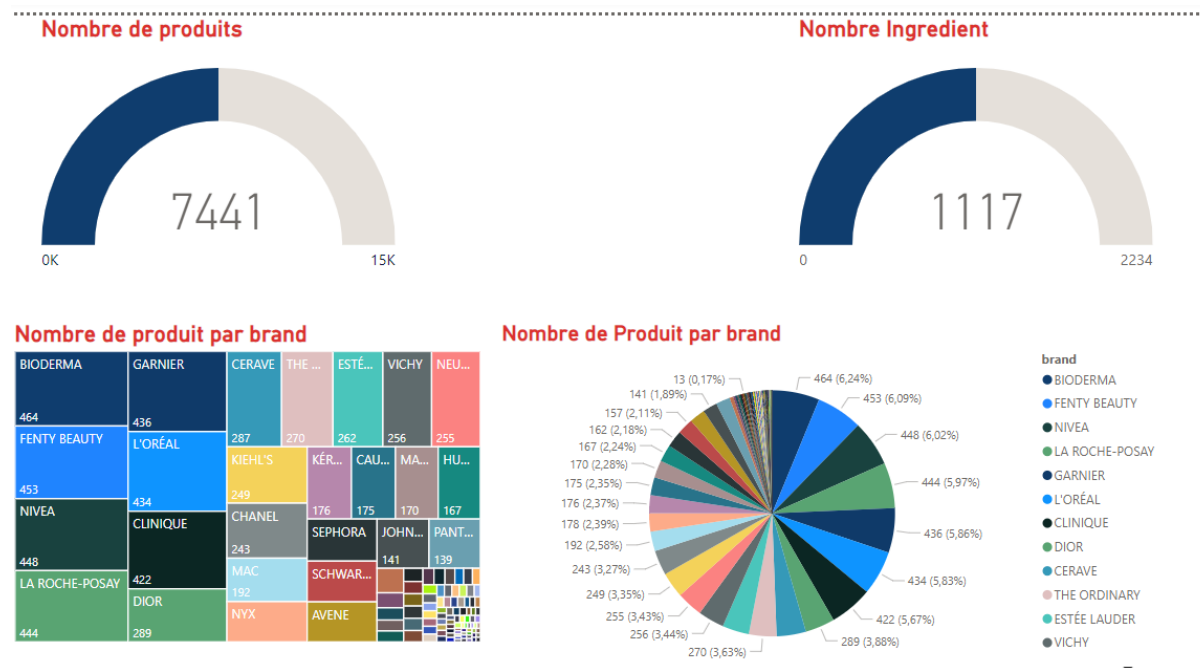
10.3 Création de la table des produits similaires

Pour analyser et regrouper les produits présentant des similitudes, nous avons mis en place une table. La méthode utilisée repose sur l'implémentation de procédures PL/SQL spécifiques étudiées en cours, notamment : Les fonctions `distance_similarity` : utilisée pour calculer les scores de similarité entre les listes d'ingrédients des produits. Cette fonction permet de quantifier le degré de ressemblance entre deux ingrédients. Ces outils ont permis de calculer précisément un score de similarité entre les produits et d'organiser les données dans une table spécifique. Cette table a été structurée pour faciliter les recherches, les analyses et les recommandations, tout en améliorant la cohérence et l'organisation des informations liées aux produits cosmétiques. Nous avons procédé par ordre alphabétique et pour chaque groupe d'ingrédient similaire, nous avons choisi un ingrédient pour les représenter.

10.4 Remplissage de la table interactions

Nous avons mis en place un script python pour nous générer tout les requêtes de mis à jour de la table interaction, à partir d'un fichier CSV qu'on a remplis par une recherche en ligne et sur chatGPT sur la toxicité d'une interaction. Cela n'a pas couvert tout les paires d'ingrédients mais juste une centaine.

10.5 Statistique



11 Perspectives

Les prochaines étapes du projet incluent l'amélioration des performances du site pour garantir une navigation fluide, ainsi que l'ajout de nouvelles fonctionnalités, telles que des filtres avancés pour affiner les recherches et des visualisations interactives permettant de mieux comprendre les relations entre les ingrédients. De plus, nous prévoyons d'enrichir la base de données avec davantage d'informations sur la toxicité des interactions, afin de fournir des analyses encore plus complètes et pertinentes. Nous envisageons également d'utiliser des algorithmes de prédiction basés sur l'apprentissage automatique ou des modèles statistiques, exploitant les propriétés chimiques et biologiques des ingrédients. Ces approches permettraient de mieux anticiper et gérer les interactions entre les ingrédients, offrant ainsi une analyse plus précise et des recommandations plus fiables.

12 Conclusion

Ce projet a permis de développer une solution innovante et fonctionnelle pour la cosmétovigilance, combinant la collecte, le traitement et la recherche de données sur les produits et ingrédients cosmétiques. Grâce à l'intégration de techniques modernes de scraping, de modélisation et d'analyse de données, nous avons établi une base de données centralisée et un outil web offrant des fonctionnalités avancées de recherche et de

recommandations. Cependant, plusieurs défis ont marqué le développement, notamment le traitement des données similaires, la gestion des doublons et l'analyse des interactions complexes entre ingrédients. Ces obstacles ont conduit à des choix techniques rigoureux et à la mise en œuvre de solutions adaptées, jetant ainsi des bases solides pour le projet. Les résultats obtenus ouvrent la voie à de nombreuses perspectives d'amélioration, telles que l'intégration de nouvelles sources de données, le perfectionnement des algorithmes de similarité, ou encore l'ajout de fonctionnalités comme des alertes personnalisées pour les utilisateurs. Avec ces évolutions, le projet pourrait non seulement renforcer la sécurité des consommateurs mais également devenir un outil incontournable dans le domaine de la cosmétovigilance.

13 Références

Cosmetify <https://www.cosmetify.com/skin-care/treatments/anti-pollution/>

BeautyBay <https://www.beautybay.com/fr/>

Site sur les acide s<https://forums.futura-sciences.com/chimie/755082-sites-acides-basiques-nucleophiles-electrophile.html>

Cosmétovigilance <https://www.vidal.fr/infos-pratiques/cosmetovigilance-et-tatouvigilance-definitions-et-modalites-de-declaration-id15205.html>

Ministère de la santé publique <https://sante.gouv.fr/soins-et-maladies/signalement-sante-gouv-fr/article/signaler-un-effet-indesirable-lie-a-un-produit-cosmetique>