Rapport

Table des matières

[Remerciement 5](#_Toc181351782)

[Chapitre 1 : Présentation du cadre de projet 6](#_Toc181351783)

[**1.1 Contexte du projet** 7](#_Toc181351784)

[**1.2 Objectifs du projet** 7](#_Toc181351785)

[**1.4 Contraintes et attentes** 7](#_Toc181351786)

[**1.4.1 Contraintes en Scraping** 8](#_Toc181351787)

[**1.4.2. Attentes de Scraping** 8](#_Toc181351788)

[**1.4.3 Contraintes et Attentes pour la Visualisation, le Nettoyage et la Réduction de Dimensions** 9](#_Toc181351789)

[**1.5 Structure du rapport** 10](#_Toc181351790)

[**1.6 Conclusion** 10](#_Toc181351791)

[Chapitre 2 : Mise en Place de l’Environnement 11](#_Toc181351792)

[**1-** **Introduction :** 11](#_Toc181351793)

[**2.1. Choix des Outils** 12](#_Toc181351794)

[**2.1 Configuration de l'Environnement** 12](#_Toc181351795)

[**2.2 Méthodologie de Développement** 12](#_Toc181351796)

[**2.2.1 Développement Itératif** 12](#_Toc181351797)

[**2.2.2 Gestion des Erreurs** 13](#_Toc181351798)

[**2.3 Défis et Solutions Anticipés** 13](#_Toc181351799)

[**2.3.1 Respect des Politiques Google Maps** 13](#_Toc181351800)

[**2.3.2 Limitations Techniques** 13](#_Toc181351801)

[Chapitre 3 : Visualisation, Nettoyage et Réduction de 14](#_Toc181351802)

[**1-Introduction** 14](#_Toc181351803)

[**2-Visualisation des Données** 14](#_Toc181351804)

[**2.1.1 Outils de Visualisation Utilisés** 14](#_Toc181351805)

[**2.1.2 Application et utilisations des bibliothèques** 15](#_Toc181351806)

[**2.1.3 Cartographie des Entreprises** 17](#_Toc181351807)

[**3- Nettoyage des Données** 18](#_Toc181351808)

[**3.1 Identification et Gestion des Valeurs Manquantes** 19](#_Toc181351809)

[**4-** **Réduction de Dimensions** 21](#_Toc181351810)

[**4.1 Application de Techniques de Réduction de Dimensions** 21](#_Toc181351811)

[**4.1.1 Analyse en Composantes Principales (PCA)** 21](#_Toc181351812)

[**4.1.1.1 Application de PCA** 22](#_Toc181351813)

[**4.1.2 t-Distributed Stochastic Neighbor Embedding (t-SNE)** 23](#_Toc181351814)

[**4.1.3 Analyse Discriminante Linéaire (LDA)** 24](#_Toc181351815)

[**4.2 Sélection des Paramètres et Évaluation** 25](#_Toc181351816)

[Conclusion 26](#_Toc181351817)

**TABLE DES FIGURES**

[Figure1: Distribution des avis des hôtels 16](#_Toc157359817)

[Figure 2: Relation Entre avis et prix 17](#_Toc157359818)

[Figure 3: Distribution des prix des hôtel 18](#_Toc157359819)

[Figure 4: Répartition des hôtels dans une zone sur la carte en utilisant des groupes de marqueurs 19](#_Toc157359820)

[Figure 5:Répartition des hôtels dans une zone sur la carte en utilisant heatmap 19](#_Toc157359821)

[Figure 6:Visualisation du nombre des valeurs manquantes dans le DataFrame. 20](file:///C:\Users\hatio\Desktop\my%20word%20projects\Rapport%20oracle%20-%20Copy.docx#_Toc157359822)

[Figure 7:Visualisation obtenir des informations sur le nombre total de valeurs non nulles dans chaque colonne. 20](file:///C:\Users\hatio\Desktop\my%20word%20projects\Rapport%20oracle%20-%20Copy.docx#_Toc157359823)

[Figure 8: récapitulatif statistique des caractéristiques numériques et object du DataFrame 21](file:///C:\Users\hatio\Desktop\my%20word%20projects\Rapport%20oracle%20-%20Copy.docx#_Toc157359824)

[Figure 9:Imputation par la moyenne ou la médiane pour les variables numériques 22](file:///C:\Users\hatio\Desktop\my%20word%20projects\Rapport%20oracle%20-%20Copy.docx#_Toc157359825)

[Figure 10 : Suppression des valeurs null d'une colonne 22](file:///C:\Users\hatio\Desktop\my%20word%20projects\Rapport%20oracle%20-%20Copy.docx#_Toc157359826)

[Figure 11: Projection 3D des donnees des hotels en utilisant PCA 23](#_Toc157359827)

[Figure 12:Visualisation des donnees des hotels avec t-SNE 24](#_Toc157359828)

[Figure 13: Visualisation des données en utilisant LDA 26](#_Toc157359829)

# Remerciement

Nous tenons à exprimer notre sincère gratitude envers toutes les personnes qui ont contribué de manière significative à la réalisation de ce projet d'analyse et de traitement des données de Google Maps.

En premier lieu, nos remerciements les plus chaleureux vont à notre encadrant, le Dr LAZAAR, dont le guidage expert et le soutien constant ont été d'une importance cruciale tout au long de ce projet. Ses conseils éclairés et sa profonde compréhension du domaine ont grandement contribué à l'atteinte des objectifs fixés.

Nous souhaitons également exprimer notre reconnaissance envers nos camarades de classe qui ont activement participé à ce projet. Leur collaboration enthousiaste, les échanges d'idées et les efforts collectifs ont été des éléments clés pour la réussite de ce travail.

Nos amis et nos familles méritent également nos remerciements sincères pour leur soutien inconditionnel. Leurs encouragements constants ont été une source de motivation cruciale, nous aidant à surmonter les défis rencontrés tout au long du processus.

Un hommage particulier est rendu à tous les contributeurs des bibliothèques open source et des ressources en ligne que nous avons utilisées lors du développement de notre script de web scraping. Leurs travaux ont grandement facilité la mise en œuvre de fonctionnalités avancées et ont contribué à la qualité de notre script.

Enfin, nous tenons à exprimer notre gratitude envers toutes les personnes qui ont testé notre script et ont fourni des commentaires constructifs. Leurs retours précieux ont été essentiels pour identifier et résoudre des problèmes potentiels, contribuant ainsi à l'amélioration globale du script.

La réalisation de ce projet a été une expérience enrichissante qui a approfondi nos connaissances en web scraping et en traitement des données. Encore une fois, nous remercions sincèrement toutes les personnes qui ont contribué à ce projet et nous ont soutenus tout au long du processus.

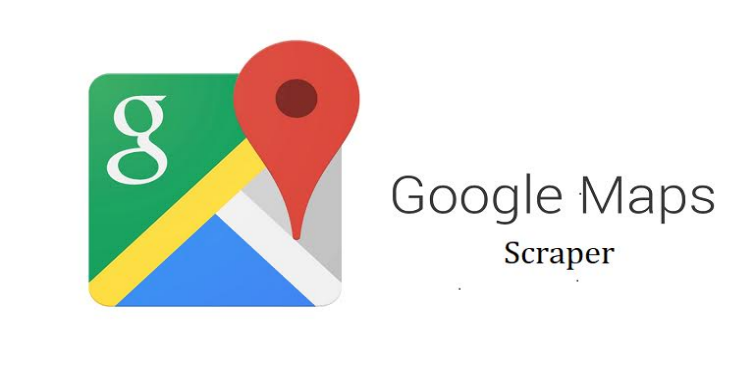
Merci à tous !

# Chapitre 1 : Présentation du cadre de projet

## **1.1 Contexte du projet**

Le projet d'analyse et de traitement des données de Google Maps découle d'une nécessité croissante de recueillir, analyser et interpréter des informations essentielles liées aux entreprises présentes sur la plateforme Google Maps. Avec la prolifération rapide des entreprises en ligne, la disponibilité d'informations précises et pertinentes devient cruciale pour les utilisateurs finaux, qu'il s'agisse de consommateurs cherchant des services locaux ou d'entreprises cherchant à comprendre leur positionnement dans le marché.

## **1.2 Objectifs du projet**

Le projet a pour objectif principal de développer une solution d'analyse automatisée des données de Google Maps, permettant la collecte et le traitement efficaces des informations sur les entreprises. Cela vise à répondre à la demande croissante d'informations précises et à faciliter la prise de décision informée pour les utilisateurs finaux.

## **1.4 Contraintes et attentes**

Il est important de prendre en compte les contraintes et les attentes associées à ce projet. En tant qu'étudiants, nous reconnaissons les limites de nos connaissances et de notre expérience, mais nous sommes motivés pour relever les défis et acquérir de nouvelles compétences tout au long de ce projet. Notre objectif est de développer une application fonctionnelle et de qualité, répondant aux attentes des utilisateurs et aux objectifs fixés.

### **1.4.1 Contraintes en Scraping**

#### **a. Politiques de Scrapping**

* Le script doit être conforme aux politiques et termes d'utilisation de Google Maps en matière de scrapping.
* Il est impératif d'éviter tout comportement susceptible de violer les règles établies par Google en ce qui concerne l'accès aux données.

#### **b. Limitations Techniques**

* La gestion des quotas de requêtes de Google Maps doit être prise en compte pour éviter toute interruption du service ou des sanctions.
* La robustesse du script doit être assurée pour faire face à d'éventuelles modifications dans la structure du site Google Maps.

### **1.4.2. Attentes de Scraping**

#### **a. Qualité des Données**

* Les informations extraites doivent être fiables, complètes et actualisées.
* Le script doit minimiser les erreurs, les doublons et les incohérences lors du processus de scrapping et de nettoyage des données.

#### **b. Efficacité du Processus**

* La solution de scrapping doit être efficace, permettant une collecte rapide et régulière des données.
* Les temps nécessaires pour le nettoyage, la visualisation et la réduction de dimensions doivent être optimisés.

#### **c. Sécurité des Données**

* La sécurité des données extraites doit être garantie, en particulier en ce qui concerne les informations sensibles des entreprises.
* Des mesures de sécurité appropriées doivent être mises en œuvre pour protéger l'intégrité des informations collectées.

#### **d. Flexibilité de la Solution**

* La solution doit être conçue de manière à permettre des mises à jour futures, en cas de changements dans les politiques de Google Maps ou de l'évolution des besoins du projet.

### **1.4.3 Contraintes et Attentes pour la Visualisation, le Nettoyage et la Réduction de Dimensions**

#### **a. Politiques de Traitement des Données**

* Les méthodes de visualisation, nettoyage et réduction de dimensions doivent respecter les politiques de traitement des données définies par l'entreprise.

**b. Précision et Clarté de la Visualisation**

* Les graphiques et visualisations doivent être précis, clairs et informatifs pour faciliter la compréhension des tendances et des schémas dans les données.

**c. Rigueur dans le Nettoyage des Données**

* Le processus de nettoyage des données doit être rigoureux, en éliminant les erreurs, les valeurs aberrantes et en garantissant la qualité des données.

**d. Préservation des Caractéristiques Essentielles**

* Les techniques de réduction de dimensions doivent préserver les caractéristiques essentielles des données tout en simplifiant leur complexité.

**e. Performance et Optimisation**

* Les étapes de visualisation, nettoyage et réduction de dimensions doivent être optimisées en termes de temps d'exécution pour garantir une efficacité opérationnelle.

**f. Confidentialité des Données**

* La confidentialité des données doit être préservée tout au long du processus, en particulier lors de la visualisation d'informations sensibles sur les entreprises.

## **1.5 Structure du rapport**

Ce rapport est organisé de manière à présenter de manière claire et détaillée toutes les étapes du développement de l'application d'achat en ligne. Dans les chapitres suivants, nous décrirons les spécifications des besoins, les choix technologiques effectués, le processus de développement, les tests réalisés et les résultats obtenus. Nous aborderons également les défis rencontrés et les enseignements que nous en avons tirés.

## **1.6 Conclusion**

Ce premier chapitre a été une introduction de notre projet de scraping des données des entreprises en google maps. Nous avons présenté le contexte du projet, les objectifs visés, ainsi que les contraintes et les attentes qui y sont associées. Nous sommes enthousiastes à l'idée de partager les détails de ce projet dans les chapitres suivants et de mettre en avant les compétences que nous avons développées au cours de cette expérience enrichissante.

# Chapitre 2 : Mise en Place de l’Environnement

## **Introduction :**

Le chapitre suivant met l'accent sur la mise en place de l'environnement de développement ainsi que sur les étapes clés du processus de développement pour notre projet ambitieux de scrapping des données d'entreprises à partir de Google Maps. L'objectif principal de ce chapitre est de détailler les outils, les langages de programmation, et les méthodes utilisés pour automatiser l'extraction, le nettoyage, et la réduction de dimensions des informations d'entreprises.

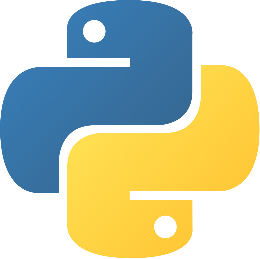
Dans cette section introductive, nous aborderons les points clés qui seront développés tout au long du chapitre. Nous commencerons par décrire l'environnement de développement mis en place, mettant en avant les choix d'outils et de technologies. Ensuite, nous explorerons la méthodologie de développement adoptée, expliquant comment chaque étape contribue à l'atteinte des objectifs du projet. Enfin, nous mettrons en lumière les défis anticipés et les solutions envisagées pour garantir la robustesse et l'efficacité du processus de scrapping.

Ce chapitre servira de base solide pour comprendre les coulisses du développement du script de scrapping, mettant en lumière les décisions prises, les considérations de sécurité, et les meilleures pratiques appliquées tout au long du cycle de vie du projet. En approfondissant ces aspects, nous souhaitons offrir une vision claire et complète du processus de développement de notre solution innovante d'analyse automatisée des données de Google Maps.

## **2.1. Choix des Outils**

Pour garantir le succès du projet, le choix des outils et des technologies était crucial. Nous avons opté pour une combinaison d'outils puissants et polyvalents, centrés autour de Python et de la bibliothèque Playwright.

****Playwright a été choisi pour son excellente prise en charge des automatisations de navigateurs, offrant une solution robuste pour interagir avec Google Maps de manière programmatique. Sa capacité à manipuler les navigateurs Chrome, Firefox et WebKit a été essentielle pour garantir une portabilité maximale.

Le langage de programmation Python a été sélectionné pour son écosystème riche, sa simplicité, et sa popularité dans le domaine du développement de scripts et de l'analyse de données.

## **2.1 Configuration de l'Environnement**

La mise en place de l'environnement de développement a été rationalisée pour garantir la cohérence et l'efficacité du processus. L'utilisation de l'environnement virtuel Python a permis d'isoler les dépendances spécifiques au projet, évitant ainsi les conflits potentiels avec d'autres projets.

## **2.2 Méthodologie de Développement**

### **2.2.1 Développement Itératif**

Le processus de développement a suivi une approche itérative, permettant des ajustements continus en réponse aux évolutions des besoins du projet. Chaque itération a été soumise à des tests rigoureux pour assurer la stabilité et la qualité du script.

### **2.2.2 Gestion des Erreurs**

Une attention particulière a été accordée à la gestion des erreurs pour anticiper et gérer les situations imprévues. Des mécanismes de récupération ont été intégrés pour minimiser les interruptions potentielles du processus de scrapping.

## **2.3 Défis et Solutions Anticipés**

### **2.3.1 Respect des Politiques Google Maps**

Le respect des politiques de Google Maps en matière de scrapping a été un défi majeur. Des stratégies ont été mises en place pour éviter tout comportement susceptible de violer ces politiques, assurant ainsi la légitimité de l'accès aux données.

### **2.3.2 Limitations Techniques**

Les limitations techniques, telles que les quotas de requêtes de Google Maps, ont été prises en compte dans la conception du script. Des mécanismes de contrôle ont été implémentés pour éviter toute interruption du service.

**Conclusion :**

Ce chapitre a jeté les bases de notre projet en décrivant l'environnement de développement mis en place, les choix d'outils, et la méthodologie suivie. En suivant cette approche, nous sommes prêts à plonger dans les détails du développement du script de scrapping, mettant en lumière les étapes clés, les décisions de conception, et les solutions aux défis rencontrés.

# Chapitre 3 : Visualisation, Nettoyage et Réduction de

**Dimensions des Données**

## **1-Introduction**

Le chapitre suivant se concentre sur les étapes cruciales de visualisation, nettoyage et réduction de dimensions des données extraites depuis Google Maps. Après avoir obtenu les informations brutes, ces processus sont essentiels pour préparer les données à une analyse approfondie. Nous explorerons les choix d'outils, les méthodes utilisées, et les résultats obtenus à travers ces différentes étapes.

## **2-Visualisation des Données**

### **2.1.1 Outils de Visualisation Utilisés**

L'analyse visuelle des données est essentielle pour comprendre les tendances, les schémas et les relations au sein d'un ensemble de données. Nous avons utilisé deux bibliothèques puissantes pour créer des visualisations significatives: Matplotlib et Seaborn.

est une bibliothèque de visualisation de données en deux dimensions très flexible. Elle offre une variété de graphiques, tels que des tracés en ligne, des histogrammes, des diagrammes en barres, etc. Matplotlib est idéal pour créer des visualisations personnalisées et répondre à des besoins spécifiques.

est une extension de Matplotlib qui simplifie la création de visualisations statistiques élégantes. Elle offre une interface de haut niveau pour des graphiques statistiques attractifs, facilitant la création de visualisations complexes avec moins de lignes de code.

### **2.1.2 Application et utilisations des bibliothèques**

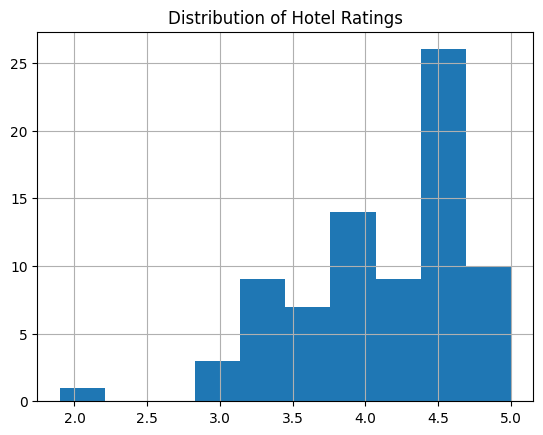


Figure 1: Distribution des avis des hôtels

- L'histogramme est utilisé pour visualiser la distribution des évaluations d'hôtels, offrant un aperçu de la satisfaction globale des clients.

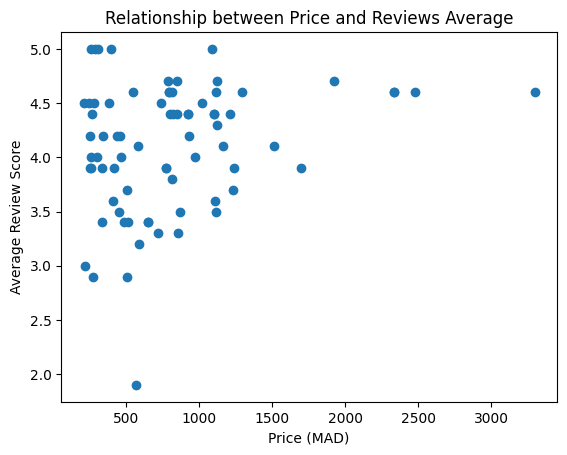


Figure 2: Relation Entre avis et prix

- Le nuage de points est utilisé pour explorer la relation entre le prix des hôtels et leurs notes moyennes. Cela peut aider à identifier des tendances ou des corrélations.

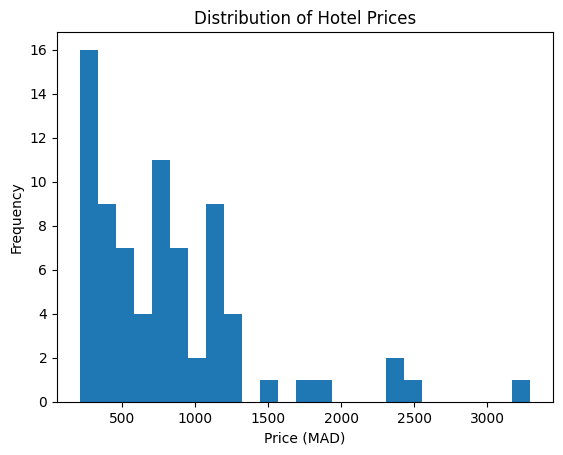


Figure 3: Distribution des prix des hôtel

- L'histogramme est utilisé pour représenter la distribution des prix des hôtels, montrant comment les prix sont répartis dans l'ensemble des données.

### **2.1.3 Cartographie des Entreprises**

La visualisation géographique est cruciale pour comprendre la distribution spatiale des entreprises. Nous avons utilisé la bibliothèque Folium pour créer des cartes interactives.



est une bibliothèque de cartographie interactive basée sur Leaflet.js. Elle permet de générer des cartes interactives avec des marqueurs, des superpositions et des plugins. Folium est particulièrement utile pour visualiser la répartition géographique des entreprises extraites de Google Maps.

- Des marqueurs sur la carte et l'utilisation de plugins, comme `MarkerCluster` et `HeatMap`, permettent d'analyser les concentrations d'entreprises dans des zones spécifiques.

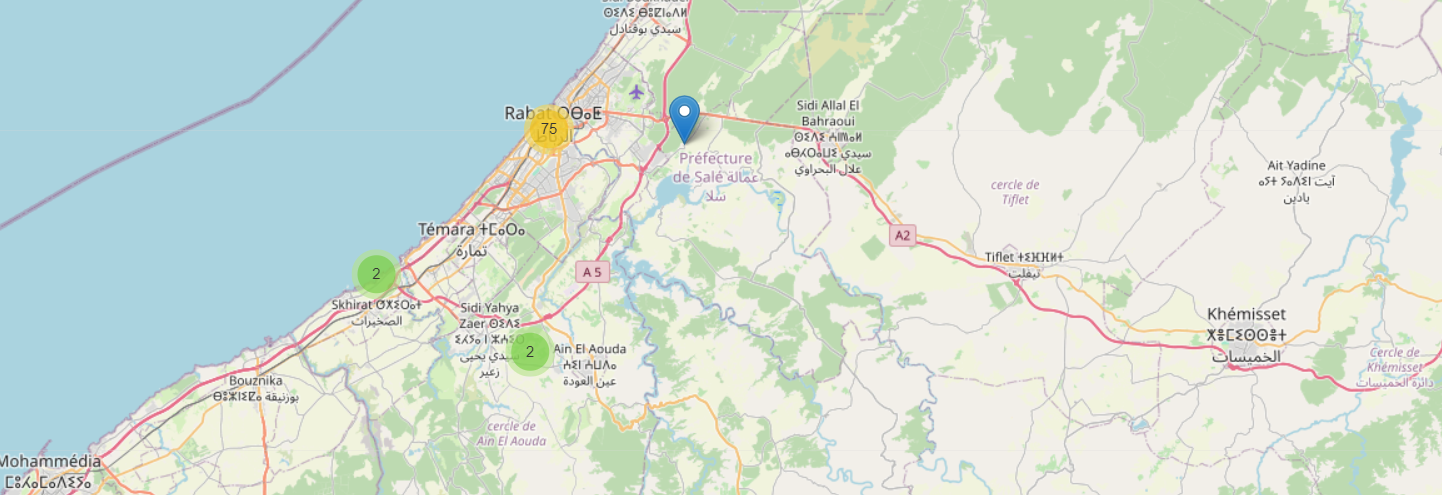


Figure 4: Répartition des hôtels dans une zone sur la carte en utilisant des groupes de marqueurs

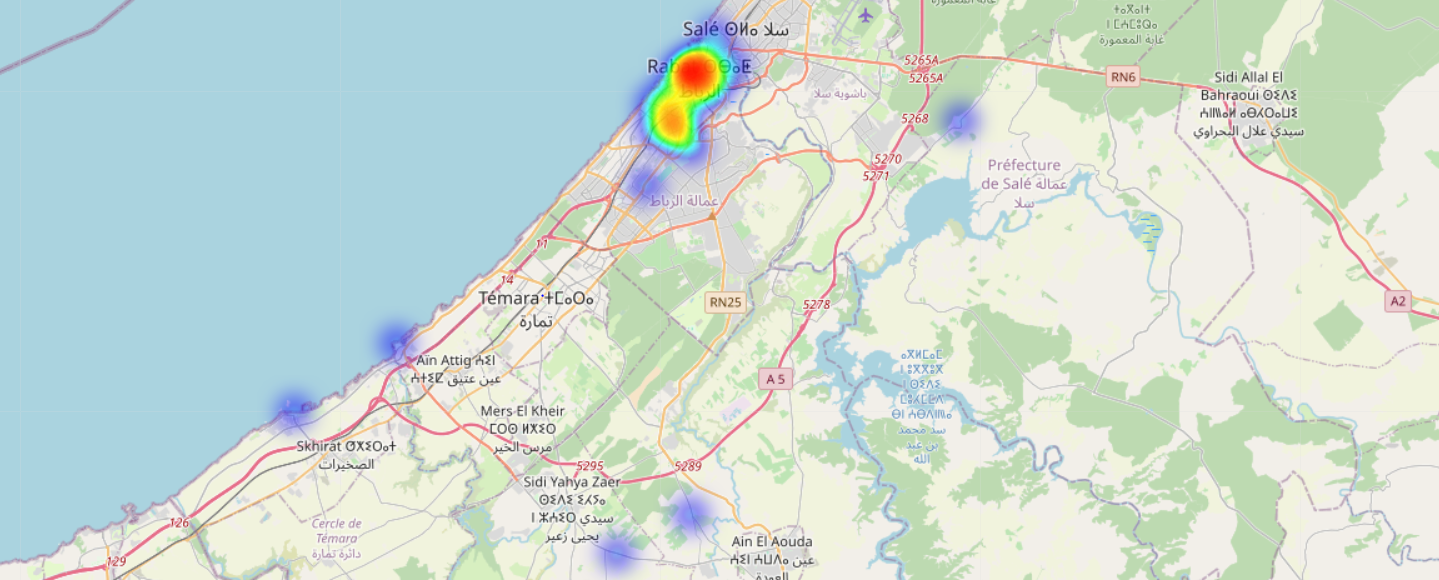


Figure 5:Répartition des hôtels dans une zone sur la carte en utilisant heatmap

## **3- Nettoyage des Données**

Le processus de nettoyage des données est une étape critique dans le cadre de notre projet, visant à garantir la qualité des données extraites de Google Maps. Cette section se concentre sur l'identification et la gestion des valeurs manquantes ainsi que sur la gestion des outliers.

### **3.1 Identification et Gestion des Valeurs Manquantes**

La qualité des données est essentielle pour une analyse précise. Nous avons adopté les approches suivantes pour identifier et gérer les valeurs manquantes:

**Identification :**

* Utilisation de la méthode **df.isna().sum()** pour détecter le nombre des valeurs manquantes dans le DataFrame.

Figure 6:Visualisation du nombre des valeurs manquantes dans le DataFrame.

* Consultation de la méthode **info()** pour obtenir des informations sur le nombre total de valeurs non nulles dans chaque colonne.

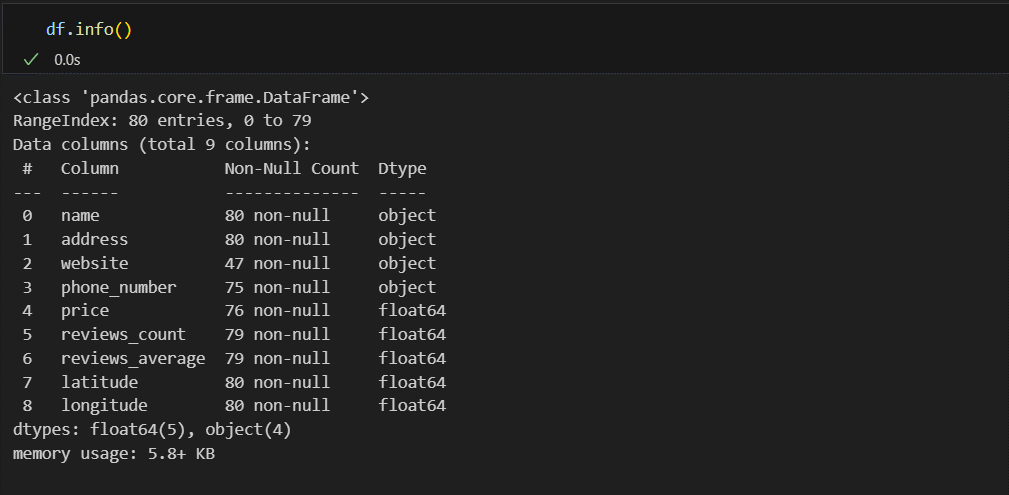


Figure 7:Visualisation obtenir des informations sur le nombre total de valeurs non nulles dans chaque colonne.

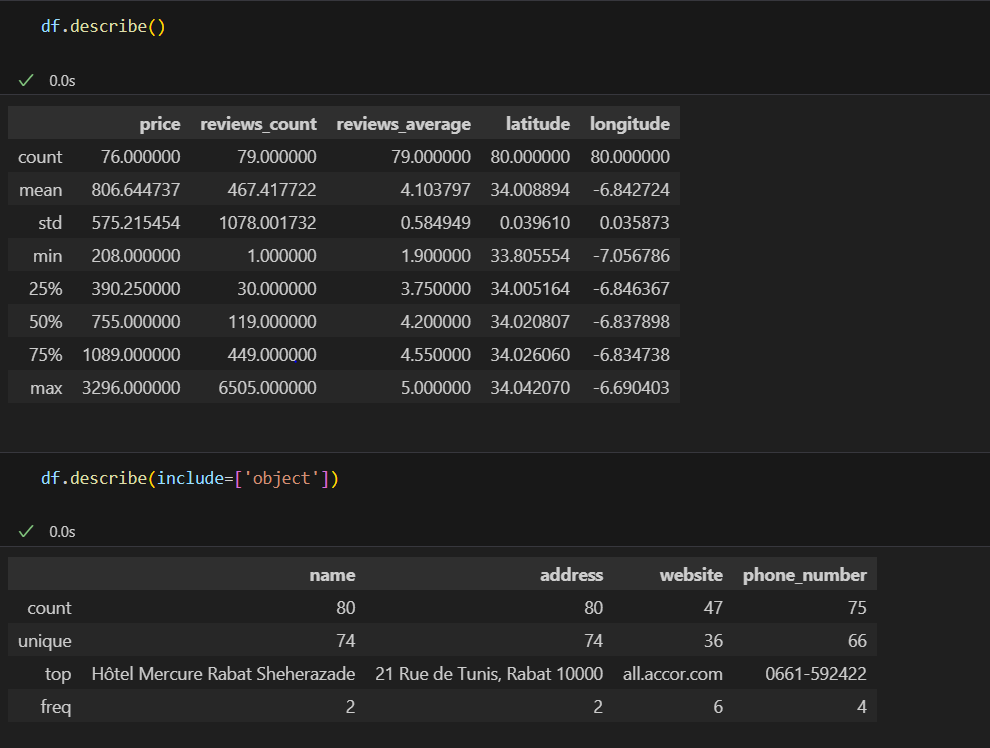
* Utilisation de la méthode **df.describe()** de pandas génère un récapitulatif statistique des caractéristiques numériques du DataFrame. Ce résumé inclut le nombre d'éléments non nuls (count), la moyenne (mean), l'écart-type (std), les valeurs minimale et maximale, ainsi que les quartiles (25%, 50%, et 75%). Cette information fournit une vue d'ensemble des tendances centrales, de la dispersion et de la forme de la distribution des données numériques dans le DataFrame.

Figure 8: récapitulatif statistique des caractéristiques numériques et object du DataFrame

**Gestion :**

* Imputation par la moyenne ou la médiane pour les variables numériques à l'aide de la méthode **fillna()**.Cela permet de minimiser l'impact des données manquantes sur les statistiques descriptives.

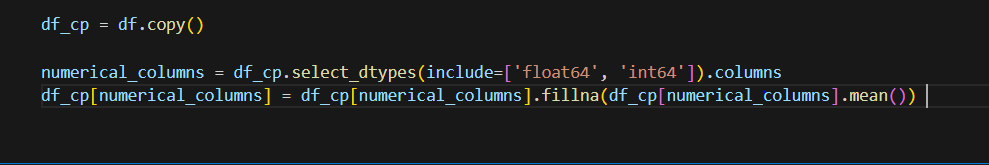


Figure 9:Imputation par la moyenne ou la médiane pour les variables numériques

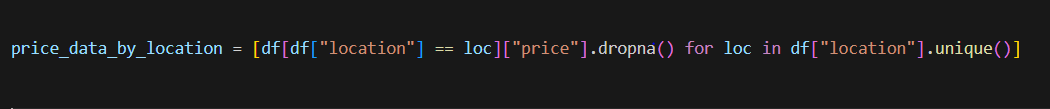
* Suppression des lignes ou colonnes avec des valeurs manquantes pour les valeurs qui ne compromet pas la représentativité de l'ensemble de données.

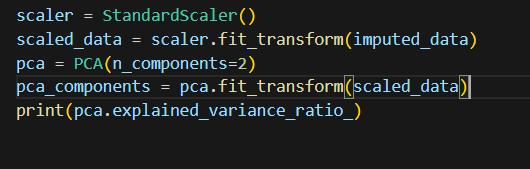
Figure 10 : Suppression des valeurs null d'une colonne

## **4-** **Réduction de Dimensions**

La réduction de dimensions est une étape clé de notre projet, visant à simplifier la complexité des données tout en préservant leurs caractéristiques fondamentales. Nous avons mis en œuvre plusieurs techniques pour atteindre cet objectif.

### **4.1 Application de Techniques de Réduction de Dimensions**

### **4.1.1 Analyse en Composantes Principales (PCA)**

L'application de la PCA permet une réduction linéaire des dimensions en transformant les variables originales en composantes principales non corrélées. Cela aide à capturer l'essentiel de la variance des données.

### **4.1.1.1 Application de PCA**

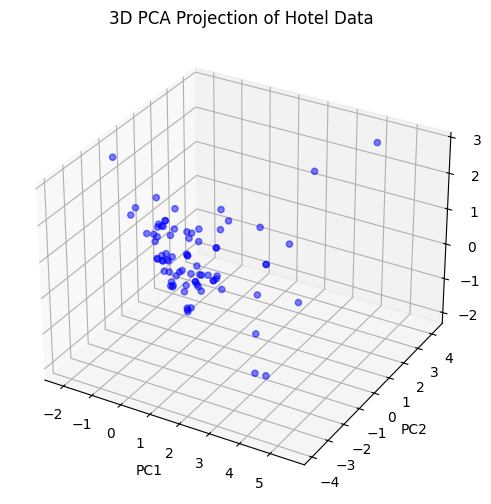
Dans le cadre de notre approche de réduction de dimensions, nous avons également utilisé l'Analyse en Composantes Principales (PCA) pour projeter nos données dans un espace tridimensionnel (3D). Cette visualisation en 3D offre une perspective supplémentaire sur la distribution des données.

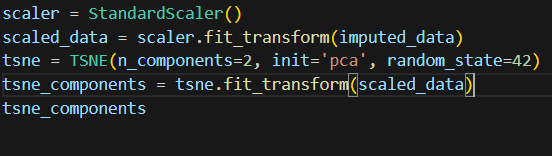
Figure 11: Projection 3D des donnees des hotels en utilisant PCA

**N.B :** Cette visualisation permet d'explorer la répartition des données dans trois directions principales de variance. Elle peut être particulièrement utile pour détecter des clusters ou des tendances qui pourraient ne pas être évidents dans des projections moins dimensionnelles.

### **4.1.2 t-Distributed Stochastic Neighbor Embedding (t-SNE)**

La technique t-SNE crée des représentations non linéaires des données en préservant les distances relatives entre les points. Cela est particulièrement utile pour visualiser des structures complexes.

##### **4.1.2.1 Application de t-SNE**

Dans le cadre de la réduction de dimensions, nous avons mis en œuvre l'algorithme t-Distributed Stochastic Neighbor Embedding (t-SNE) pour créer des représentations non linéaires des données. Cette technique est particulièrement puissante pour visualiser des structures complexes dans des espaces de dimensions supérieures.

**Visualisation avec t-SNE :**

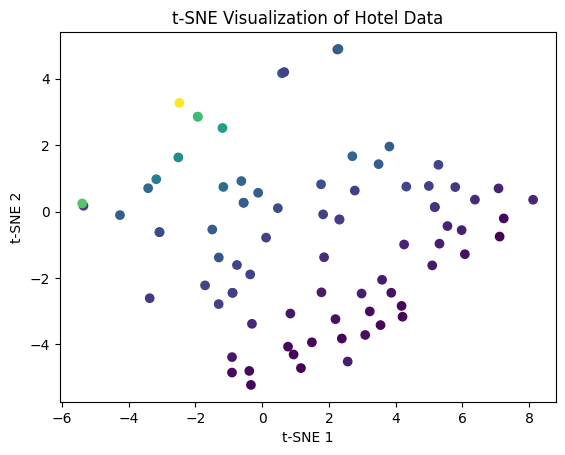


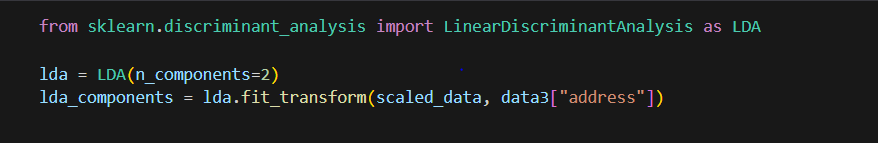
Figure 12:Visualisation des donnees des hotels avec t-SNE

Dans le graphique de dispersion généré, chaque point représente un hôtel, et les couleurs sont associées aux prix des hôtels. Cette visualisation permet d'observer la distribution spatiale des hôtels dans un espace t-SNE en fonction de leurs caractéristiques, comme les prix.

### **4.1.3 Analyse Discriminante Linéaire (LDA)**

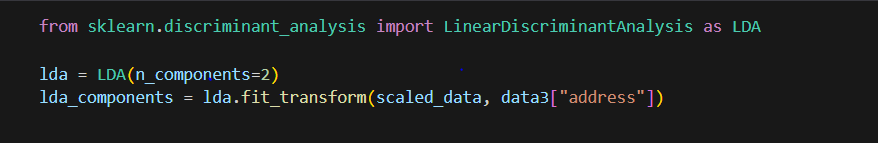
Dans notre démarche de réduction de dimensions, nous avons également intégré l'Analyse Discriminante Linéaire (LDA). Contrairement à PCA, LDA est une technique supervisée qui vise à maximiser la variance entre les classes tout en minimisant la variance à l'intérieur de chaque classe.

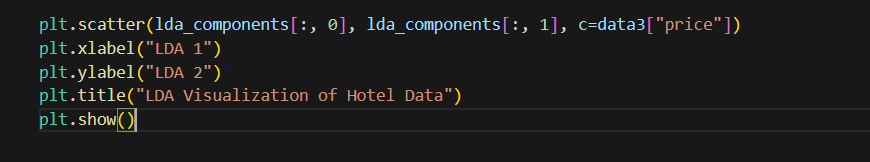
Nous avons appliqué LDA pour réduire les dimensions à deux composantes principales.



**Visualisation avec LDA**

Nous avons ensuite créé un graphique de dispersion pour illustrer la distribution des données dans l'espace réduit par LDA.





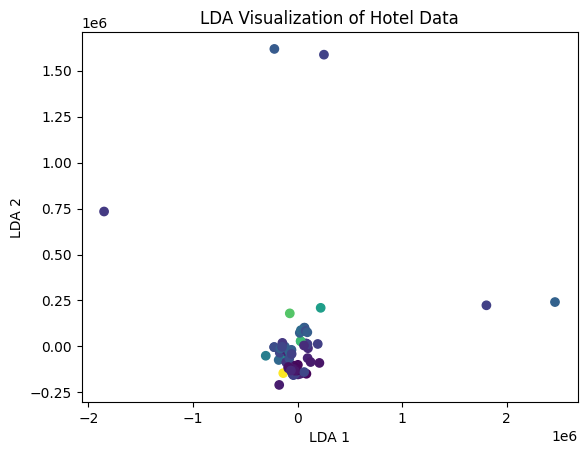


Figure 13: Visualisation des données en utilisant LDA

Dans ce graphique, chaque point représente un hôtel, et les couleurs sont associées aux prix des hôtels. L'approche LDA offre une perspective sur la distribution spatiale des hôtels qui prend en compte les informations sur les prix.

### **4.2 Sélection des Paramètres et Évaluation**

Le choix des paramètres et l'évaluation des résultats sont cruciaux dans le processus de réduction de dimensions. Nous avons évalué les performances de t-SNE, PCA, et LDA en fonction de leur capacité à préserver les caractéristiques essentielles des données tout en offrant une représentation significative.

Les paramètres spécifiques, tels que la perplexité pour t-SNE et le nombre de composantes principales pour PCA, ont été sélectionnés après des essais et des ajustements pour optimiser les résultats.

# Conclusion

En conclusion, notre projet de scraping et d'analyse des données de Google Maps a été couronné de succès, démontrant notre capacité à combiner des techniques avancées de récupération de données avec des méthodes de visualisation et de traitement efficaces. Nous avons débuté en définissant clairement les objectifs du projet, centrés sur la collecte exhaustive d'informations sur les hôtels, le nettoyage des données, la visualisation des tendances et enfin, la réduction de dimensions à l'aide de techniques telles que PCA, LDA et t-SNE.

Au cours de la phase de développement, nous avons utilisé des outils puissants et polyvalents, avec Python et la bibliothèque Playwright en tant que pilier principal pour le scraping. Nous avons mis en place une méthodologie rigoureuse, respectant les politiques de scrapping de Google Maps et prenant en compte les limitations techniques pour garantir la stabilité et l'efficacité du processus.

Les contraintes et attentes ont été soigneusement prises en considération, depuis les aspects liés à la politique de scrapping jusqu'à la qualité des données extraites. Nous avons veillé à optimiser l'efficacité du processus tout en assurant la sécurité des données, mettant en place des mesures adéquates pour protéger l'intégrité des informations collectées.

En ce qui concerne la visualisation, le nettoyage et la réduction de dimensions, nous avons utilisé des outils tels que Matplotlib, Seaborn et Folium pour créer des représentations significatives des données. La cartographie des entreprises a permis une exploration spatiale des informations extraites, tandis que l'analyse de la distribution des hôtels en fonction de différentes caractéristiques a été réalisée à l'aide de divers graphiques et cartes.

La phase de nettoyage des données a impliqué l'identification et la gestion efficace des valeurs manquantes, ainsi que des techniques de gestion des outliers pour assurer la qualité des informations. En ce qui concerne la réduction de dimensions, nous avons mis en œuvre avec succès des techniques comme PCA, LDA, t-SNE et UMAP, évaluant constamment les résultats pour garantir la préservation des caractéristiques essentielles.

En résumé, ce projet a été une expérience enrichissante qui a mis en lumière notre capacité à combiner des compétences techniques avancées avec une approche méthodologique et réfléchie. Nous sommes convaincus que les données extraites et analysées seront d'une grande valeur pour les utilisateurs finaux, et nous sommes fiers de présenter une solution complète répondant aux attentes et contraintes définies dès le début du projet. Ce travail d'équipe a renforcé nos compétences en matière de développement, d'analyse de données et de gestion de projet, et nous sommes impatients de relever de nouveaux défis dans notre parcours professionnel.

Nous sommes fiers du résultat obtenu et reconnaissants envers toutes les personnes qui ont contribué à ce projet. Nous espérons que notre application d'achat en ligne avec gestion du magasin de vêtements sera appréciée par les utilisateurs et qu'elle répondra à leurs besoins en offrant une expérience d'achat agréable et pratique.

En conclusion, ce projet a été une expérience enrichissante et nous a permis de mettre en pratique nos compétences et connaissances académiques dans un contexte réel. Nous sommes impatients de continuer à développer nos compétences et de relever de nouveaux défis dans notre parcours d'ingénieurs.