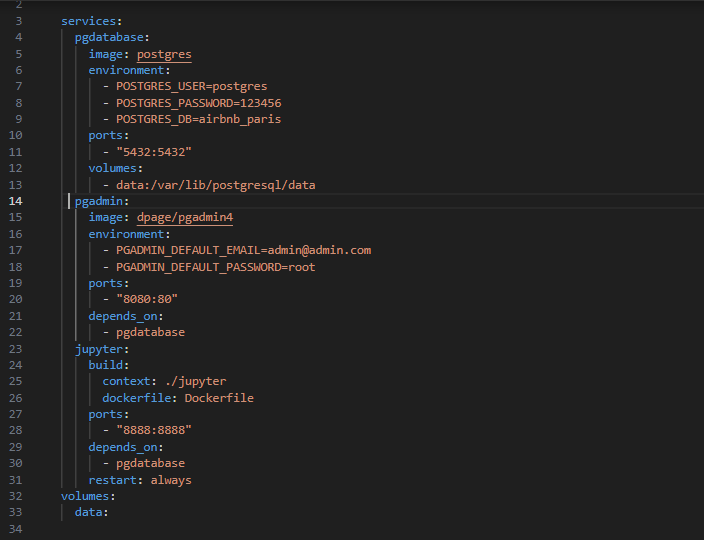
### **Introduction**

Le projet "Utilis Libres" vise à fournir une plateforme robuste pour l'analyse de données sur les hébergements Airbnb à Paris, en anticipation des Jeux Olympiques de 2024. Cette plateforme s'appuie sur un ensemble de services conteneurisés, spécifiquement configurés pour faciliter l'analyse de données et la gestion de bases de données.

docker: c est un utilitaire on l utiliser our facilite developpement on a utiliser 3 service et on contenerisé avec le docker compose



#### 1. Service pgdatabase (PostgreSQL)

Description :

Le service pgdatabase utilise l'image Docker officielle de PostgreSQL. Il est configuré pour stocker les données concernant les hébergements Airbnb à Paris.

Configuration :

* Image : postgres
* Variables d'environnement :
  + POSTGRES\_USER: postgres
  + POSTGRES\_PASSWORD: 123456
  + POSTGRES\_DB: airbnb\_paris
* Ports :
  + Exposé : 5432
* Volumes :
  + Persiste les données dans un volume nommé data.

Rôle :

Ce service sert de base de données principale pour le stockage et la gestion des données Airbnb.

#### 2. Service pgadmin (pgAdmin 4)

Description :

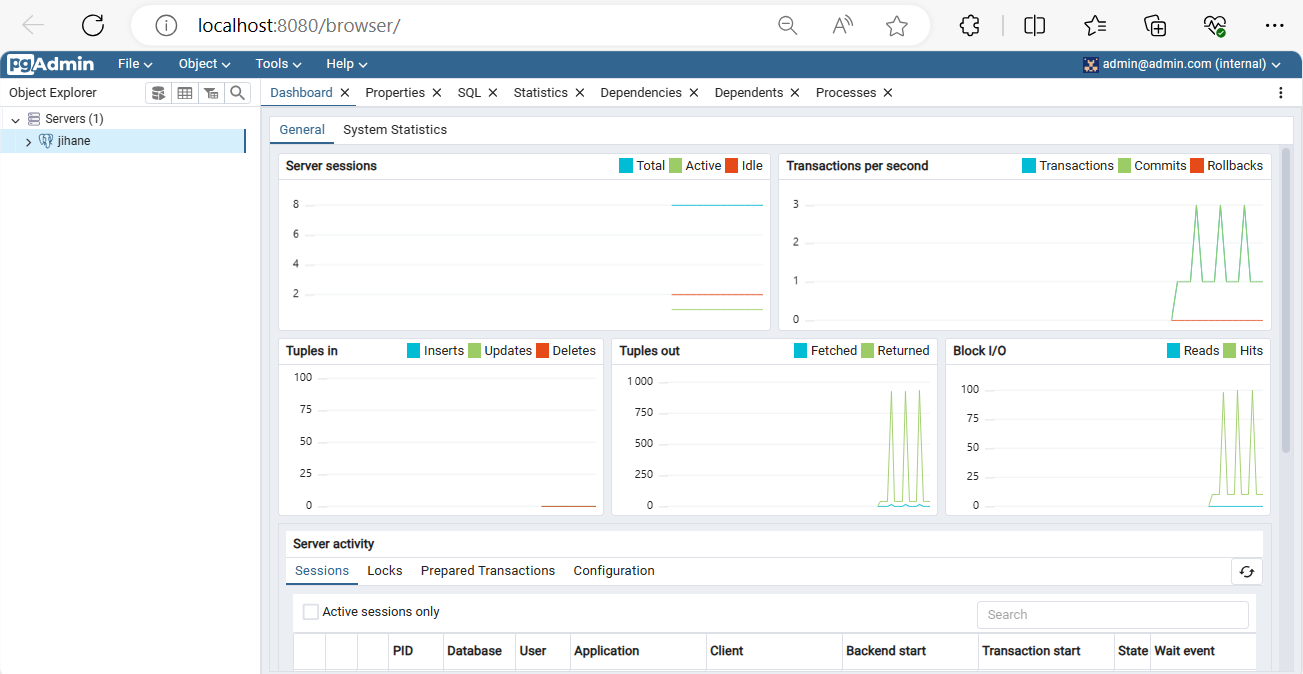
pgadmin fournit une interface web pour la gestion de la base de données PostgreSQL. Il facilite les requêtes, la visualisation et la gestion des données.

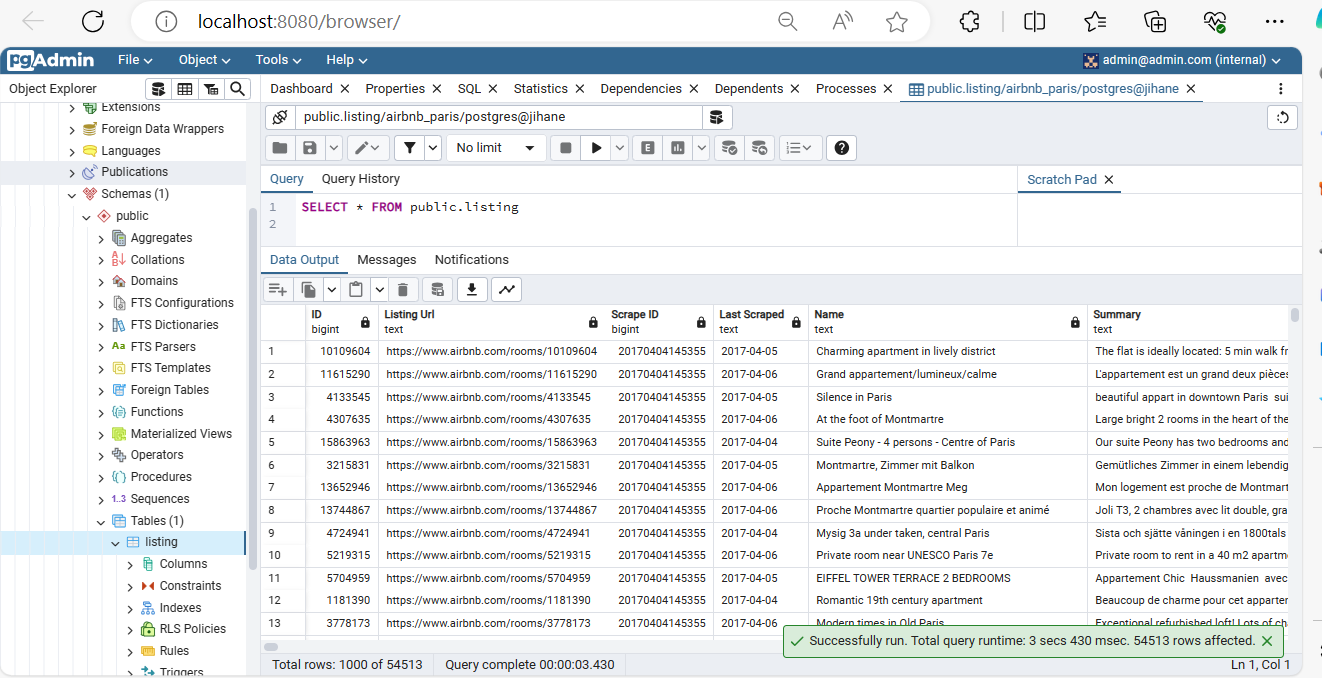
Configuration :

* Image : dpage/pgadmin4
* Variables d'environnement :
  + PGADMIN\_DEFAULT\_EMAIL: admin@admin.com
  + PGADMIN\_DEFAULT\_PASSWORD: root
* Ports :
  + Exposé : 8080
* Dépendances :
  + Dépend du service pgdatabase.

Rôle :

Ce service permet aux analystes de données d'interagir facilement avec la base de données PostgreSQL à travers une interface graphique.





#### 3. Service jupyter (Jupyter Notebook)

Description :

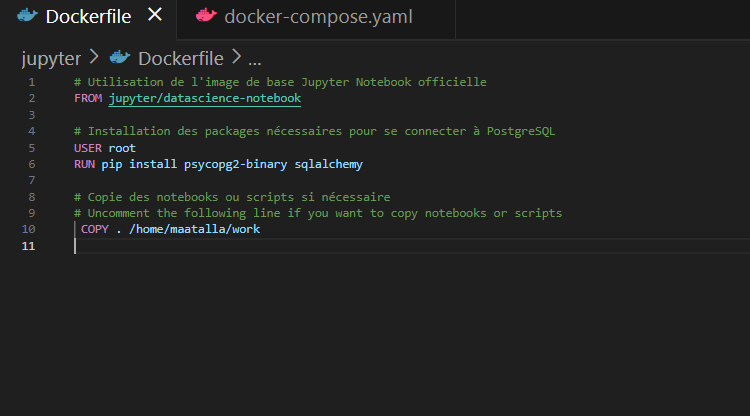
Le service jupyter fournit un environnement Jupyter Notebook, permettant l'exécution de code Python, la visualisation de données, et l'analyse statistique.

Configuration :

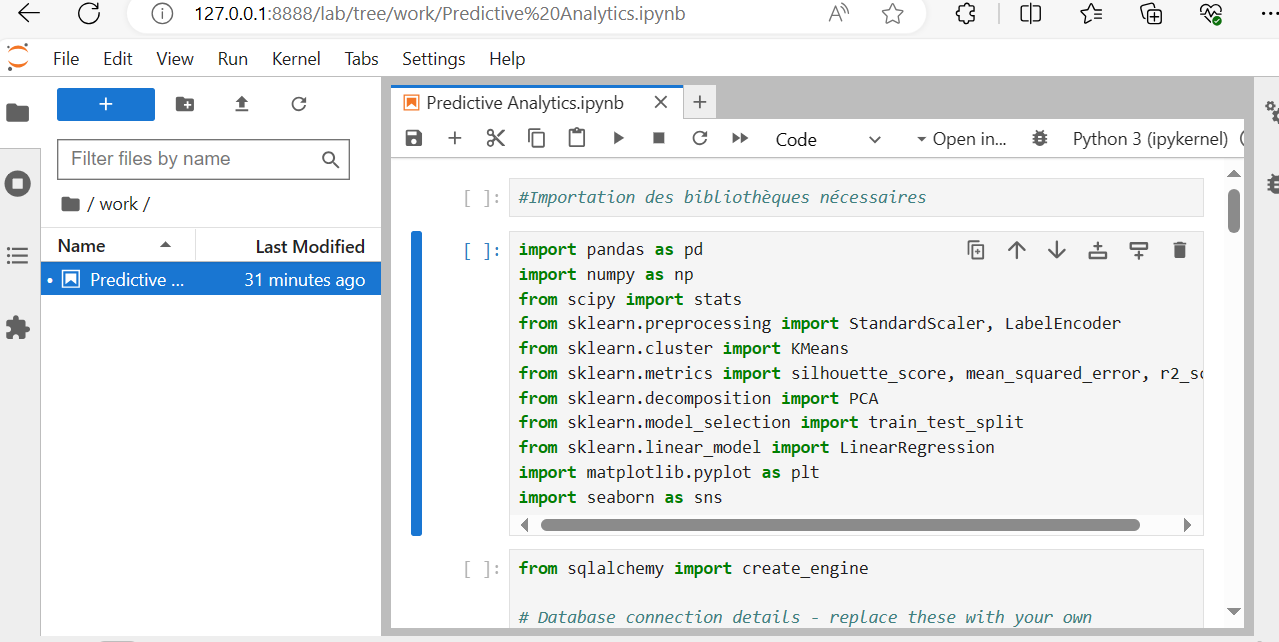
* Build :
  + Contexte : ./jupyter
  + Dockerfile : Dockerfile spécifié dans le répertoire jupyter.
* Ports :
  + Exposé : 8888
* Dépendances :
  + Dépend du service pgdatabase.
* Redémarrage : automatique en cas d'arrêt inopiné.

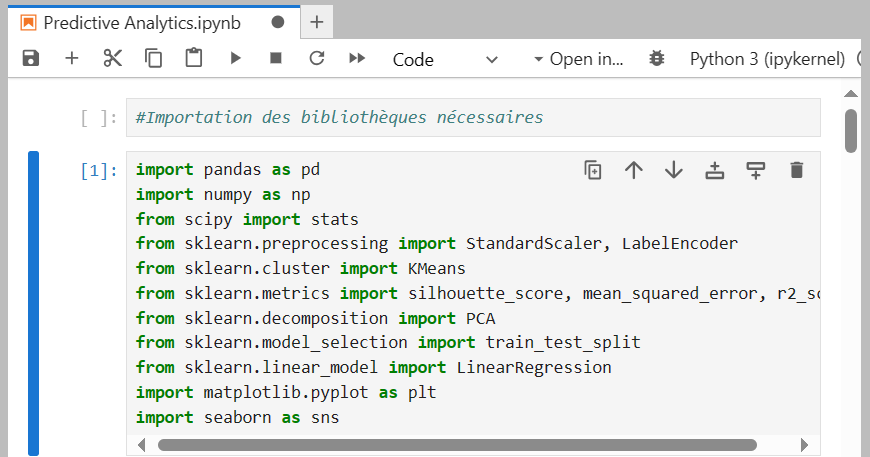
Rôle :

Jupyter est un outil essentiel pour les analystes, offrant un environnement interactif pour l'exploration de données, le développement de modèles prédictifs, et la visualisation des résultats.

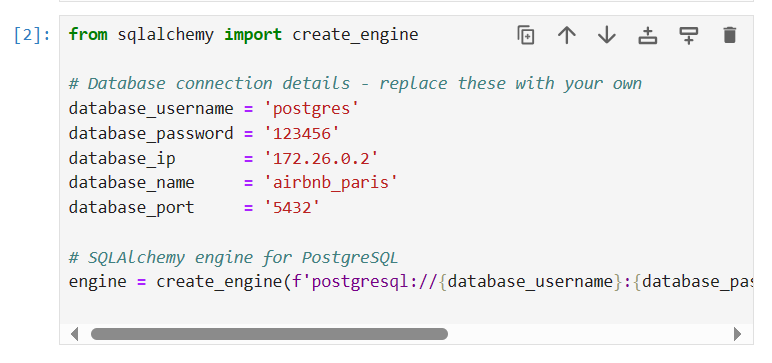


chargement du fichier notebook



import tion des bibliothéques 

importer bibliothéque sqlalchemy pour communiquer avec postgress



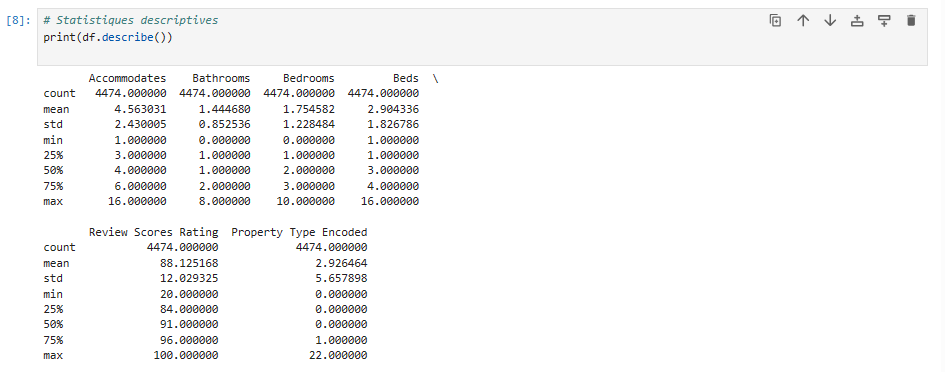
supprimer les colonnes et les lignes dupliquer

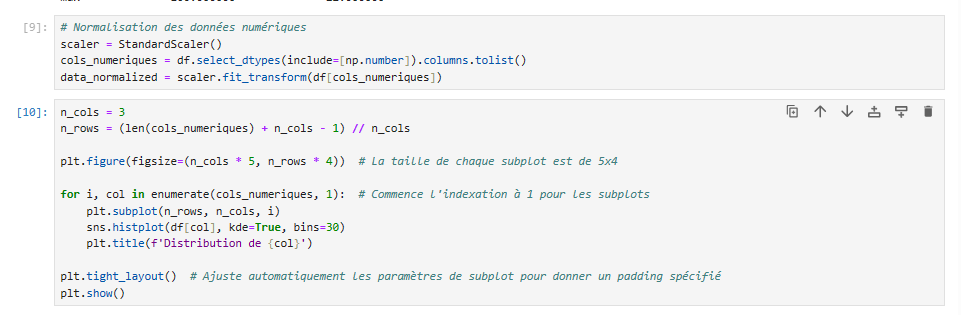


supprimer les valeurs null 

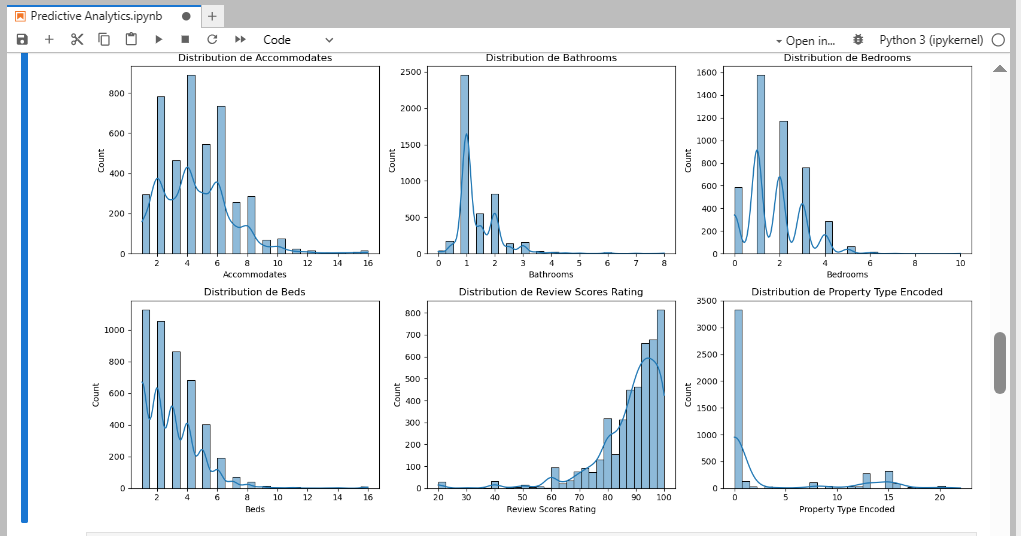
changer le type de colonne en numérique



afficher les statiqtiques 

normalisé et afficher la distribution des données 

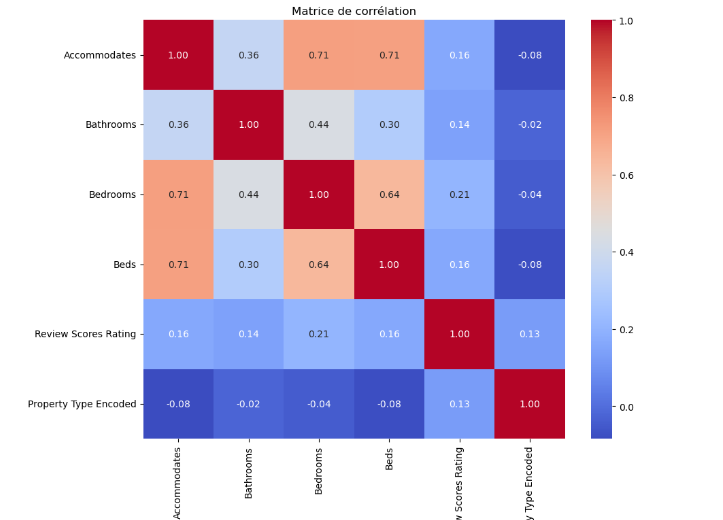
vesualiser la distribution des données



afficher la matrice de corelation



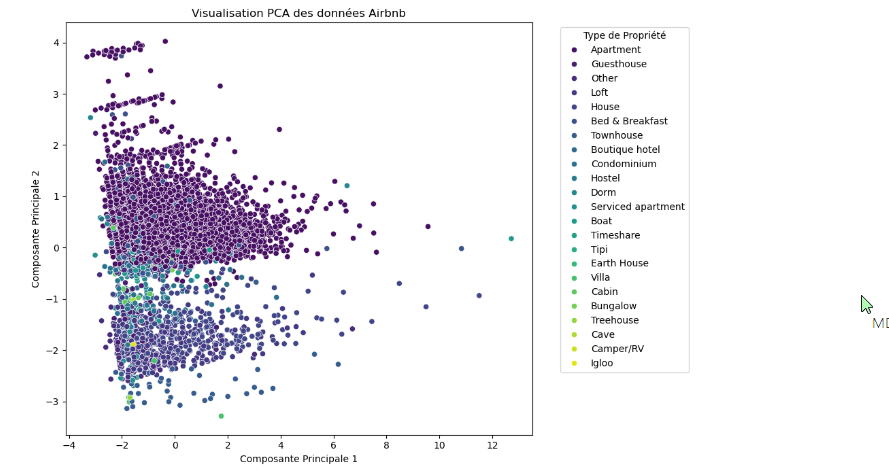
visualiser la matrice de colorration



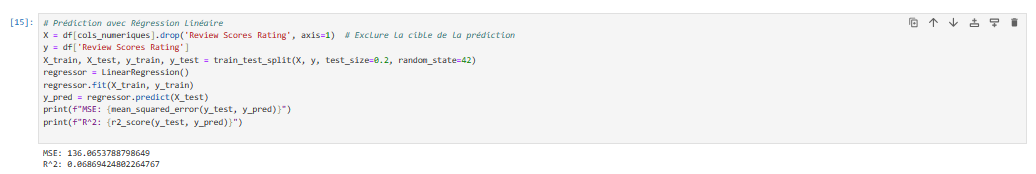
Reduction et dimension avec ACP



visualisation PCA des données AIRBNB



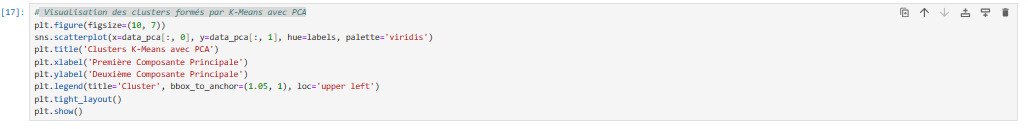
Prédiction avec Régression Linéaire

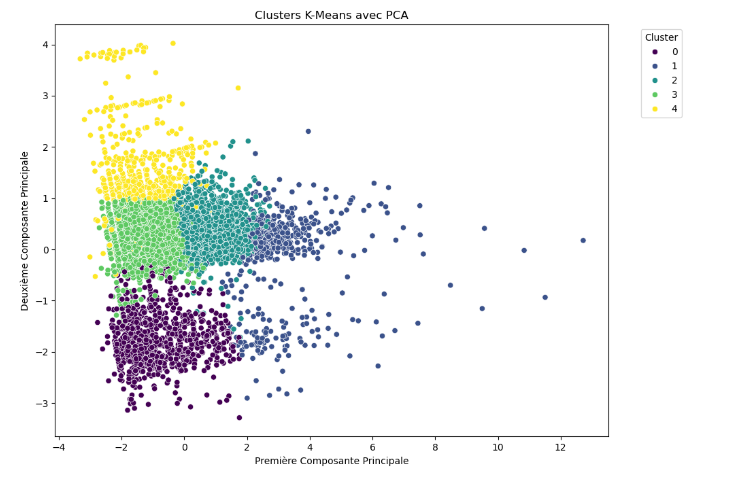


clustring avec kmeans



Visualisation des clusters formés par K-Means avec PCA





il reste conclusion et motivation

**Github**