

Étapes d'analyse et de visualisation avec Matplotlib

1. Télécharger le dataset

Téléchargez le dataset depuis Kaggle en utilisant le lien Pima Indians Diabetes Dataset. Après l'avoir téléchargé, vous pourrez l'utiliser localement pour vos analyses.

2. Importer les bibliothèques nécessaires

Importez les bibliothèques nécessaires pour le projet :

- Pandas : pour la manipulation des données.
- Matplotlib : pour créer des graphiques.

3. Charger et explorer le dataset

- a. Informations générales sur le dataset : Chargez les données à partir du fichier CSV, affichez les premières lignes et examinez les types de données.
- b. Statistiques descriptives : Obtenez des statistiques descriptives pour les colonnes numériques (moyenne, écart-type, valeurs minimales, maximales, etc.).
- c. Grouper par une catégorie : Groupez les données par la colonne 'Outcome' et calculez des statistiques pour chaque groupe.
- d. Trier les groupes : Triez les groupes par une variable comme le niveau de glucose pour mieux comparer les statistiques entre les groupes.

4. Prétraiter les données

Vérifiez les valeurs manquantes et traitez-les en les remplissant par la moyenne ou en les supprimant si nécessaire.

5. Créer des graphiques avec Matplotlib

- a. Histogramme : Créez un histogramme pour visualiser la distribution d'une variable comme l'âge des patients.

- b. Diagramme à barres : Créez un diagramme à barres pour visualiser le nombre de patients diabétiques et non diabétiques.
- c. Diagramme circulaire (Pie chart) : Créez un diagramme circulaire pour visualiser la proportion de patients diabétiques.
- d. Diagramme à moustaches (Boxplot) : Créez un boxplot pour visualiser la distribution des niveaux de glucose.

6. Afficher plusieurs graphiques en un seul coup d'œil

Affichez plusieurs graphiques sur une seule figure en utilisant des sous-graphes pour une comparaison plus facile.

7. Graphique en nuage de points (Scatter Plot)

Créez un nuage de points pour visualiser la relation entre deux variables comme l'âge et le niveau de glucose.

8. Bubble Chart

Créez un graphique à bulles pour montrer trois dimensions de données en même temps, comme l'âge, le glucose et la pression artérielle.

9. Graphique en lignes

Créez un graphique en lignes pour visualiser la tendance d'une variable, par exemple l'évolution du niveau moyen de glucose par âge.

10. Sauvegarder vos graphiques

Sauvegardez un de vos graphiques générés sous forme d'images (par exemple PNG ou JPEG) pour une utilisation ultérieure.