

Mini Projet : Analyse de données

1) Objectif du Mini projet :

L'objectif de ce mini projet est de réaliser une analyse de données sur un data set, afin d'explorer, d'identifier et de comprendre les relations, les motifs et les tendances présents dans un ensemble de données. Cette analyse vise à extraire des informations sur les variables étudiées, leurs interactions et leurs influences mutuelles. En les réalisant de manière séquentielle et progressive, elle permet de créer une vue complète et approfondie des données, facilitant ainsi la prise de décision.

Pour bien comprendre et interpréter les données, il est souvent recommandé de suivre les étapes d'analyse univariée, bivariée et multivariée de manière progressive. Ces étapes permettent d'explorer et d'interpréter les différentes facettes des données, en partant des analyses plus simples pour arriver à des analyses plus complexes et détaillées. Ces étapes d'analyse univariée, bivariée et multivariée sont généralement importantes et souvent recommandées pour réaliser un projet d'analyse de données de manière méthodique et approfondie.

Pour ce faire, l'étudiant doit utiliser le langage Python pour analyser les données, d'un data set spécifique. Il devra mettre en œuvre les trois catégories d'analyses (univariée, bivariée et multivariée) et ensuite élaborer un rapport approfondi qui expose la méthodologie adoptée pour ces analyses.

2) Le dataset

Vous avez besoin d'un ensemble de données initial, également appelé un jeu de données. L'ensemble de données est la base sur laquelle vous effectuerez toutes les étapes d'analyse, qu'il s'agisse d'analyses univariées, bivariées, ou multivariées.

Un ensemble de données peut être constitué de diverses observations et de leurs attributs ou variables. Ces variables peuvent être de différents types : numériques (continues ou discrètes) ou catégorielles.

Travail demandé :

1. Décrire le dataset utilisé, (lien de téléchargement, nombre d'attribut, nombre d'individus, type de variable ...) vous pouvez télécharger à partir des portails des données tels que : UCI , Kaggle ...
2. L'objectif à réaliser sur le dataset.
3. Prétraitement du dataset : Il est important de prétraiter et organiser vos données de manière appropriée avant de commencer l'analyse. Le processus de prétraitement peut inclure la gestion des valeurs manquantes, la déduplication, la normalisation des données, la transformation si nécessaire, et bien d'autres étapes pour s'assurer que les données sont cohérentes et prêtes à être analysées.

3) Analyse Univariée :

L'objectif principal de l'analyse univariée est de comprendre la distribution et les caractéristiques d'une seule variable à la fois. Cette analyse implique des méthodes statistiques et graphiques pour examiner les tendances centrales, les mesures de dispersion, ... etc. L'analyse univariée permet d'obtenir un aperçu initial de chaque variable et de détecter des informations potentiellement importantes.

Travail demandé :

1. Commencer par analyser chaque variable individuellement.
2. Calculer les statistiques descriptives telles que la moyenne, la médiane, l'écart type, etc.
3. Créer des graphiques appropriés comme des histogrammes, des diagrammes en boîte, ou des graphiques de densité pour visualiser la distribution des données. Cela vous donne une idée de la nature de chaque variable et de ses caractéristiques.

4) Analyse Bivariée :

L'analyse bivariée se concentre sur la relation entre deux variables à la fois. L'objectif est de déterminer s'il existe des relations ou des associations significatives entre les paires de variables. Cela peut aider à identifier des corrélations, des dépendances ou des différences statistiquement significatives entre les variables. L'analyse bivariée permet d'explorer comment une variable peut être influencée par une autre.

Travail demandé :

Une fois que vous avez compris chaque variable individuellement,

1. Explorer les relations entre les paires de variables.
2. Utiliser des tests statistiques appropriés (comme le test de corrélation ou le test du chi-carré) peuvent être utilisés pour évaluer la force et la signification de ces relations.
3. Utiliser des graphiques les plus appropriés pour visualiser les associations entre deux variables quantitatives ou qualitatives, vous pouvez utiliser des tableaux croisés et des graphiques en barres pour identifier les tendances entre différentes catégories.
4. Interpréter les résultats trouvés.

5) Analyse Multivariée :

L'analyse multivariée implique l'examen simultané de trois variables ou plus. Son objectif est de comprendre les relations complexes entre plusieurs variables et de déterminer comment elles interagissent les unes avec les autres. Les techniques d'analyse multivariée peuvent inclure l'analyse, l'analyse factorielle, l'analyse en composantes principales et d'autres méthodes qui permettent d'explorer les relations multidimensionnelles entre les variables.

Travail demandé :

Une fois que vous avez identifié des relations potentielles entre les paires de variables, passez à l'analyse multivariée pour explorer les interactions plus complexes entre trois variables ou plus. Vous pouvez utiliser des techniques telles que :

1. L'ACP ou l'AFC, ou autres méthodes qui sont utiles pour réduire la dimensionnalité des données et identifier des motifs cachés.
2. Utiliser les graphiques nécessaires à la visualisation des résultats.
3. Interpréter les résultats trouvés.

6) Rédaction du rapport :

Rédiger un rapport décrivant l'ensemble des étapes citées contenant la description du dataset, et les différents types d'analyse, pour obtenir une vue complète et approfondie des données étudiées. Avec le détail de chaque type d'analyse.