

Projet

Niveau : L3

Modalités

Le travail est à faire consiste à manipuler les fichiers csv, les structures de données (dataframe, matrices, listes, ...). Le travail est à faire par groupe de 3 étudiants. Chaque groupe devra écrire les fonctions en Python pour répondre aux différentes questions et un programme principal qui utilise ces fonctions. A la fin, chaque groupe devra préparer un exposé de 5 minutes pour présenter le travail effectué.

Description des données

Le fichier « employees.csv » décrit les employés d'un commerce avec les attributs numéro (employeeNumber), nom de famille (lastName), prénom (firstName), son extension (extension), son adresse email (email), son bureau (officeCode), son supérieur (reportsTo) et son poste (jobTitle). Le fichier « clients » contient les données sur les clients décrites par les attributs suivants : numéro (customerNumber), son nom (customerName), son contact (contactLastName), son nom (contactFirstName), son téléphone (phone), sa première adresse (addressLine1), sa deuxième adresse (addressLine2), son ville (city), son état (state), son code postal (postalCode), son pays (country), un numéro de l'employé en charge (salesRepEmployeeNumber), un montant de credit limite (creditLimit)

TAF

Prétraitement

Ces deux fichiers contiennent les données manquantes (lignes vides, valeurs NULL). Il faut les traiter.

- 1) Écrire les fonctions Python pour créer des nouveaux fichiers « client_new.csv » et « employees_new.csv » à partir des fichiers décrits plus haut de telle sorte que les lignes vides soient supprimées
- 2) écrire une fonction qui remplace le code postal ayant pour valeur NULL, remplacer par « 00000 »

Statistiques

écrire des fonctions pour :

1. afficher les statistiques d'une colonne de données. Elle prend comme paramètre la nom de la colonne (ou le numéro) et un numéro (0 pour client et 1 pour employees) indiquant le client, puis pour les attributs numériques affiche les statistiques (min, max moyenne, nombre de valeurs manquants) et une courbe pour visualiser ces valeurs. Pour les attributs non numériques, la fonction devra afficher le nombre de valeurs manquantes, pour chaque valeur sa fréquence ; et enfin un histogramme pour visualiser.
2. Pour un employé (dont le numéro est passé en paramètre) le nombre d'employés dont il est le supérieur. Utiliser cette fonction pour afficher cette donnée pour tous les employés
3. Pour un employé (dont le numéro est passé en paramètre) le nombre de clients dont il a la charge. Utiliser cette fonction pour afficher cette données pour tous les employés
4. Proposer des histogrammes pour visualiser ces données

Croisement des données

1. écrire des fonctions qui prend comme paramètre un numéro de l'employé et retourne la somme des crédits des clients dont il a la chargé.
2. Écrire une fonction qui crée un fichier comportant pour chaque employé le montant total des crédits de tous les clients à sa charge. Utiliser la fonction précédente
3. Écrire une fonction qui crée un fichier comportant pour chaque ville le montant total des crédits de tous les clients qui y sont basés.
4. Écrire une fonction qui crée un fichier comportant pour chaque pays le montant total des crédits de tous les clients qui y sont basés.
5. Pour tous les résultats obtenus, proposer des outils graphiques pour les visualiser