

Projet de traitement de données - Cahier des charges

Groupe :

- Laurène VILLACAMPA
- Ludovic DENEUVILLE
- Jean-Philippe TROTTA

Partie 1 : Présentation du projet

Ce travail s'inscrit dans le cadre du cours d'introduction à la programmation objet. Le but du projet est de réaliser un programme permettant d'analyser des données. Dans notre cadre d'étude, il s'agit de mettre en place un processus de traitement de fichiers de données. En sortie de notre programme, l'utilisateur obtiendra des données propres qu'il pourra utiliser pour des traitements statistiques. Dans un premier temps, les données utilisées sont les relevés météorologiques et les relevés de consommation électrique en France et sur les dix dernières années.

Nous disposons de deux jeux de données :

- Jeu de données issu du site de Météo France :
Ce jeu de données recense des données relatives à la météo mesurées dans les stations françaises (hors Corse) entre janvier 2023 et mars 2022. Il contient un fichier .csv par mois :
 - Maille géographique : station
 - Pas temporel : 3 Heures
 - Identifiants : instant (Date-heure) et station (ID)
- Jeu de données issu du site de Réseaux Energie :
Ce jeu de données donne la consommation quotidienne brute régionale en électricité, exprimée en MW, en France (hors Corse) sur la période allant de janvier 2013 à décembre(!) 2022. Il contient un fichier .json par mois :
 - Maille géographique : région
 - Pas temporel : 1/2 Heure

- Identifiants : instant (Date-heure) et région (code INSEE region)

Le programme issu de ce travail doit donc implémenter des fonctionnalités permettant l'analyse des données proposées. Il doit en outre pouvoir être réutilisé assez facilement, et être aisément adaptable à de nouvelles utilisations.

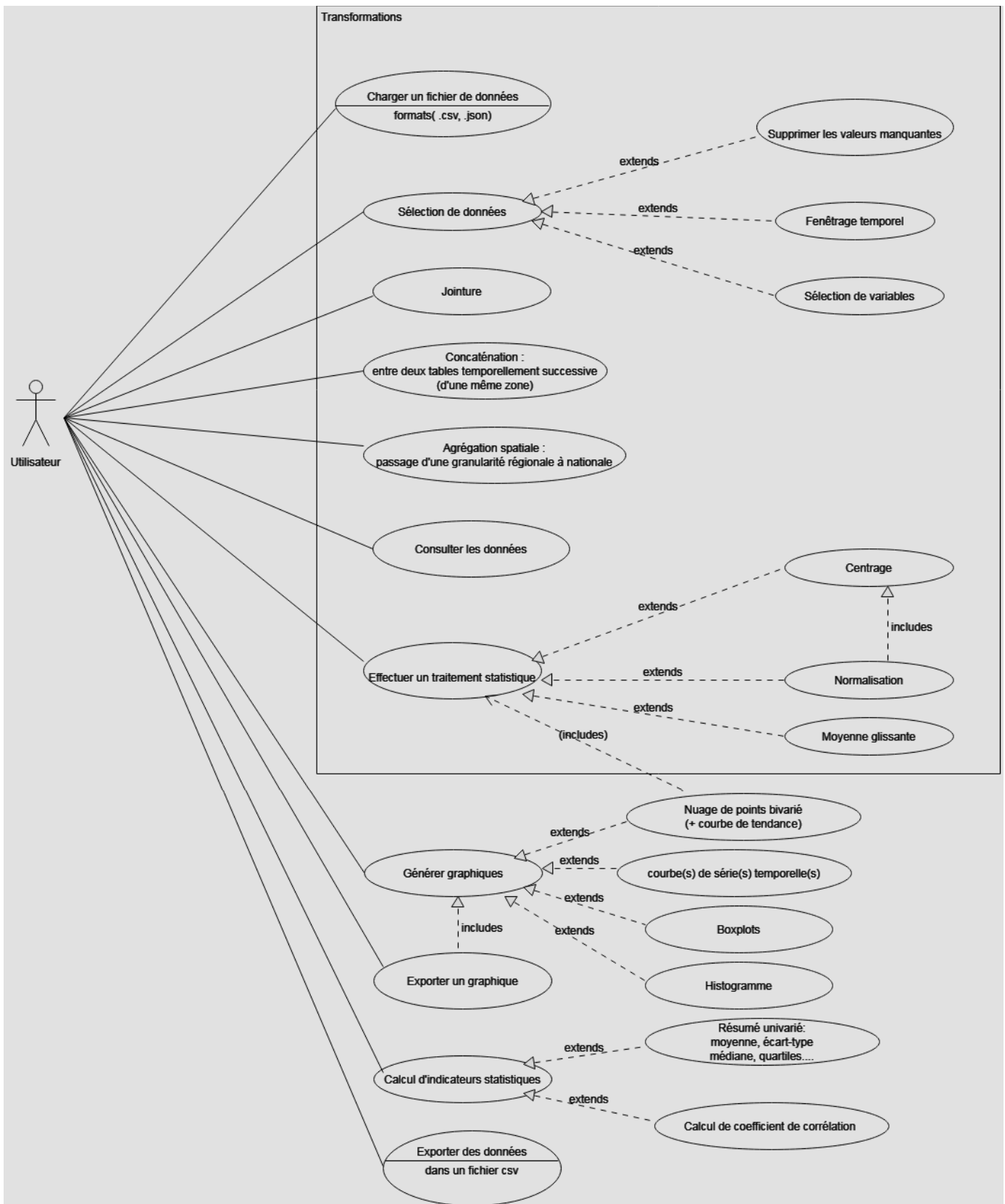
Partie 2 : Contraintes

Plusieurs contraintes techniques sont imposées pour ce projet.

- Le travail demandé doit être réalisé dans le langage Python. Certains packages sont autorisés, à savoir `numpy`, `matplotlib` et `scipy`. A contrario, l'utilisation du package `pandas` est interdite. D'autres packages peuvent éventuellement être utilisés sur autorisation. Le module `datetime` qui fait partie de Python peut être utilisé.
- Dans une logique de programmation orientée objet, le programme doit être modulaire, réutilisable et évolutif. Toutes les classes devront être documentées et les méthodes testées.

Partie 3 : Résultats attendus

Les utilisations attendues de ce programme sont résumées dans le diagramme de cas d'utilisation ci-dessous :



Partie 4 : Description des fonctionnalités

Nous allons détailler ici les principales fonctionnalités définies dans le diagramme de cas d'utilisation.

Charger un fichier de données

- L'utilisateur clique sur le bouton **Charger un fichier**
- Une fenêtre s'ouvre et permet de sélectionner un fichier (un seul) présent sur la machine
 - pour l'instant, seuls les formats csv et json sont acceptés
- L'utilisateur doit donner le nom de la table dans laquelle sera chargée le fichier
- Une confirmation est demandée : "Voulez vous charger le fichier xxx ?"
 - Si oui le programme lance le chargement du fichier
 - s'il y a une erreur, un message explicite est affiché
 - sinon les données sont chargées dans la base de données

Exemple : la table donnees_meteo a été créée

Sélectionner des données

- L'utilisateur clique sur le bouton **Sélectionner des données**
- Il sélectionne les variables qu'il souhaite
 - pour chaque variable, il peut fenêtrer, c'est à dire
 - choisir une valeur minimale et/ou maximale pour les variables quantitatives
 - pour les variables qualitatives, choisir parmi une liste
 - il peut choisir une liste de variables de la table pour lesquelles il souhaite éliminer les individus (lignes) pour lesquels il y a au moins une valeur manquante
 - il peut choisir des variables provenant de différentes tables en réalisant d'abord une jointure (cf ci-dessous)
- Il précise le nom de la nouvelle table qu'il crée
- Il clique sur `Créer la nouvelle table`

Exemple : la table meteo_electricite_dep35 a été créée

Gestion des données

- L'utilisateur clique sur **Gérer les données**, il arrive sur un écran listant les tables
- Pour chacune des tables, il peut :

- Consulter la table, c'est à dire afficher les données sous forme de tableau
- Supprimer la table
- Il y a également un bouton **Faire une jointure entre deux tables**
 - Sélectionner les 2 tables, puis sélectionner les colonnes à joindre
- et enfin un bouton **Concaténer des tables**
 - l'utilisateur sélectionne les tables à concaténer et donne le nom de la nouvelle table
 - le programme vérifie que toutes les variables de chaque tables sont au même format
 - en cas d'erreur, un message explicite s'affiche, sinon la nouvelle table est créée

Effectuer un traitement statistique

- L'utilisateur clique sur le bouton **Effectuer un traitement statistique**
- Il sélectionne une table puis une variable
- Il sélectionne le traitement qu'il veut opérer sur cette variable :
 - centrage, normalisation, demander une moyenne glissante
- Il valide et cela crée une nouvelle variable

Exemple : la variable temperature a été centrée

Exporter une table

- L'utilisateur clique sur le bouton **Exporter une table**
- Il sélectionne la table qu'il veut exporter et choisit le nom du fichier
- Il choisit le format d'export (pour le moment seul le format csv est disponible) et clique sur le bouton Exporter

Exemple : la table meteo35 a été exportée dans le fichier meteo35.csv

Afficher des graphiques

- L'utilisateur clique sur le bouton **Afficher un graphique**
- Il peut ensuite sélectionner le type de graphique
 - pour le moment, un seul type sera disponible : nuage de points
- Il sélectionne ensuite 2 variables
 - éventuellement un fenêtrage sur ces variables
- Il clique sur le bouton Générer graphique et le graphique s'affiche
 - il y a un bouton retour arrière pour revenir à la sélection des variables

