
German Credit Data Set

Le TP précédent vous a permis d'aborder les différentes étapes de la fouille de données. Il vous a également permis d'apprendre à manipuler le logiciel SAS Enterprise Miner à travers son tutoriel. Cette séance vous permettra de revoir ces notions sur un autre exemple guidé et commenté.

L'utilisation du module Enterprise Miner nécessite quelques connaissances de la base SAS sur laquelle il s'appuie. Cette séance a également pour objectif de connaître les différents éléments de l'interface de la base SAS et de comprendre l'organisation du logiciel ainsi que la gestion des fichiers (Exercices 1 et 2).

Exercice 1 : L'interface du logiciel SAS.

Ouvrez le logiciel SAS sans ouvrir Enterprise Miner. Découvrez chacun des éléments suivants :

- une barre d'outils, située en haut de l'interface,
- une barre de navigation entre les différentes fenêtres ouvertes du logiciel, située en bas de l'interface (utilisez cette barre pour visualiser les fenêtres cachées par d'autres),
- une fenêtre d'explorateur de fichiers, situé à gauche de l'interface,
- une fenêtre d'accès aux résultats issus des programmes SAS, située derrière la fenêtre d'explorateur,
- une fenêtre pour éditer les programmes, située au dessus de la barre de navigation,
- une fenêtre de log utilisée par le logiciel pour afficher des messages à chaque action de celui-ci, elle est située au dessus de la fenêtre d'édition de programmes,
- une fenêtre pour l'affichage des données issues de programmes SAS, située derrière la fenêtre de log.

A partir du menu d'aide de la barre d'outils, ouvrez la documentation de la base SAS en cliquant sur "Introduction au système SAS", puis cliquez sur "Go". Différentes pages de documentation sont proposées. Par exemple, cliquez sur "Explore the SAS workspace" pour obtenir des informations sur les différents éléments de l'interface du logiciel. Cette documentation de la base SAS ne sera pas distribuée, n'hésitez pas à vous y référer.

Exercice 2 : Gestion des fichiers de données dans SAS.

SAS possède ses propres bibliothèques de fichiers de données. Ces bibliothèques sont des répertoires où les fichiers de données sont stockés. Un des objectifs de cet exercice est de vous créer votre propre bibliothèque qui vous servira, durant toute la durée de l'enseignement, à stocker les jeux de données sur lesquels vous allez travailler.

Ouvrez le tutoriel de la base SAS comme indiqué à l'exercice précédent : par l'intermédiaire du menu d'aide de la barre d'outils. Ouvrez la section "Work with SAS Data sets" et suivez le tutoriel jusqu'à sa fin (une dizaine de pages). Ce tutoriel va vous guider à effectuer les étapes suivantes :

- découvrir les bibliothèques SAS et visualiser leurs fichiers,
- créer votre propre bibliothèque,
- déposer un fichier de données dans cette bibliothèque et le renommer,
- ouvrir un fichier de données,
- voir les propriétés de ce jeu de données,
- voir la liste des attributs des données,
- modifier le format d'un de ces attributs,
- visualiser la liste des formats disponibles dans SAS.

Enterprise Miner utilise les bibliothèques que vous venez de découvrir. Les jeux de données "Dining List" et "German Credit Data Set" sont inclus d'origine dans ces bibliothèques, les prochains ne le seront pas, c'est la raison pour laquelle vous devez connaître leur fonctionnement.

Exercice 3 : Le German Credit Data Set.

A partir du tutoriel de SAS que vous avez déjà ouvert, trouvez "Example Process Flow Diagram" en cliquant successivement sur "SAS Products", "Enterprise Miner" et "Getting Started With SAS Enterprise" dans le menu à gauche de la fenêtre. Ce deuxième tutoriel d'Enterprise Miner vous propose un autre exemple de fouille de données sur le jeu de données "German Credit Data Set". Suivez ce tutoriel et effectuez les étapes indiquées simultanément.