Automatiser les traitements va devenir rapidement indispensable. On ne peut pas à chaque fois que l'on a des choses à faire, les analyser à la main ; il faut un moment donné se rendre compte que le volume des données et leur diversité, ne serait-ce qu'à travers le temps, nécessitent une automatisation des traitements ; ce qui n'empêche que c'est bien l'homme qui va réaliser l'automatisation de ces traitements grâce à des méthodes qui vont permettre de faire parler ces données. D'où la nécessité d'apprendre un petit peu de base de programmation et en ce moment le langage qui a le vent en poupe et qui est idéal pour analyser les données, c'est Python.

Alors dans ma formation il n'est pas question de faire de vous des spécialistes de Python mais de vous montrer qu'il n'est pas difficile de prendre en main ce langage et d'effectuer déjà des traitements pour analyser de façon fine un jeu de données (réaliser quelques petits traitements qui font donc partie de ce que j'appelle l'automatisation de l'analyse des données).

Il vous sera ensuite loisible de pousser plus loin votre formation du langage Python afin de réaliser des traitements de plus en plus sophistiqués de vos données qui vous permettront d'aller encore plus loin dans l'analyse de ces données et de parvenir à des modèles ou des prédictions de plus en plus précis qui permettront effectivement de faire parler ces fameuses données de façon efficace et juste. Cependant, n'oubliez pas que le plus important ce sont les données et non les outils qui permettent de les analyser car ces outils ne révèlent aucune surprise. Autrement dit, c'est le couple données/méthode choisie pour faire parler vos données qui va être le plus important. L'outil n'étant que le moyen simple d'arriver rapidement à un résultat et de l'automatiser. Alors pour être sûr de ces données, il faut être sûr de leurs sources, de la façon dont elles ont été mesurées et collectées, maîtriser les erreurs et des biais qui peuvent être inhérents à cette collecte et traiter également les éventuelles données manquantes et cetera et cetera. Certaines fonctions mathématiques et statistiques de base seront suffisantes pour analyser les données. Il faudra parfois peut-être pousser vers des fonctions et des méthodes statistiques un peu plus élaborées pour obtenir des modèles qui pourront intégrer de plus en plus de facteurs pour expliquer ce que l'on observe et pour avoir des prédictions plus fiables à partir des données que l'on a collectées ou dont on dispose.

Nous allons prendre deux exemples pour initier notre formation au *machine learning* avec le langage Python. Le premier va être simplement une prise en main du langage qui va nous permettre d'analyser le fichier du suivi des arbres (jeu de données) à l'aide des fonctions Python de base et de la bibliothèque Pandas. Le deuxième exemple va consister à se servir d'un fichier de données pour prédire la nature de fruits et de légumes à partir de descripteurs et autres paramètres qui sont enregistrés dans ce fichier, sans intervention humaine. Ce genre d'applications existe déjà dans certains magasins ou supermarchés qui détectent automatiquement les fruits et légumes et, à partir du poids qui est enregistré soit dans le caddie soit dans la chaîne, on obtient le prix de ce qu'il faut payer.

Il ne me reste plus qu'à vous dire à bientôt en classe pour passer à la pratique et surtout n'hésitez pas à intervenir pour demander des explications sur tel ou tel aspect de la formation ou pour orienter cette formation vers un aspect auquel je n'aurais pas pensé.