Formation Python : Data science (2 jours)

L’essor de la data dans les entreprises a propulsé Python en têtes des langages les plus utilisés. Le besoin de traitement et de visualisation de données est devenu indispensable. Python propose des outils très puissant et facile d’utilisation dans le traitement des données : Numpy, Pandas… et la visualisation de données : Matplotlib, Seaborn…

Cette formation est destinée à un public ayant les bases de Python. Elle intéresse les data scientist, les data analyst, les scientifiques et encore simple utilisateur d’Excel. Cette formation apprend à récupérer des données automatiquement, à les traiter et à les restituer dans des visuels. Les outils de traitement et visualisation de données sont approfondis. Enfin elle pose les briques fondamentales du machine-learning.

Jour 1

1. Anaconda
   * Qu’est-ce qu’un environnement data ?
   * Présentation de la plateforme Anaconda, de ses outils Jupyter notebook et Spyder.
2. Numpy
   * Les différentes structures de données (ndarray, matrice)
   * Intégration de données de fichiers « .csv »
   * Manipulation de données ( Agréger, insérer, supprimer, filtrer, slicing…)
   * Restructurer un tableau (Fractionner, combiner, fusionner)
   * Calcul matriciel
   * Résolution d’équation
   * Calcul statistique (moyenne, médiane et variance)
   * Les tableaux masqués
3. Pandas
   * Créer un DataFrame
   * Intégration de données de fichiers Excel
   * Indexing avec les méthodes « loc() » et « iloc() »
   * Manipulation de données (Filtrer, trier, supprimer, remplacer, extraire, créer des colonnes…)
   * Restructurer un tableau (Fractionner, combiner, fusionner, pivoter, dépivoter)
   * Décrire les données avec les outils statistiques

Jour 2

1. Matplotlib
   * Récupérer les données
   * Créer un visuel (courbe, diagramme, histogramme, secteur, nuage de points et visualisation 3D)
   * Mise en forme d’un visuel
2. Seaborn
   * Récupérer les données
   * Les différents types de visuels : relation, distribution, catégorisation, corrélation, régression et multi-visuel.
   * Création d’un visuel : Lineplot, scatterplot, displot, hisplot, kdeplot, catplot, boxplot, heatmap…
3. SciPy
   * Interpolation
   * Calcul matriciel
   * Régression linéaire
4. Beautiful Soup
   * Généralités sur le web-scraping
   * Chargement d’une page WEB
   * Naviguer dans la structure de la page WEB
   * Rechercher des éléments