STAT4DECISION
conseil et formation data science

FORMATION PYTHON POUR LA DATA SCIENCE

REF. : PYTHON101

DESCRIPTION

Trois jours intensifs pour acquérir les bases de l'utilisation du langage python pour des applications en analyse de données, data mining et data science.

Après une introduction au langage python (ou un rappel si vous avez déjà utilisé python), vous apprendrez à maîtriser les bibliothèques spécialisées en data science, data mining et data visualisation. Basée sur de nombreux exercices pratiques, cette formation vous permet d'être immédiatement opérationnel pour le traitement de vos données grâce à python.

OBJECTIFS

- Maîtriser les bases de python
- Etre capable d'analyser des données avec python
- Etre capable d'automatiser le traitement de données avec python
- Connaître les outils pour la data science de python

DUREE

3 jours

PUBLIC ET PREREQUIS

Tout public ayant des bases en traitement de données.

MOYENS PEDAGOGIQUES

Alternance d'exposés et d'applications pratiques avec des exercices sur des données.

PLUS D'INFORMATIONS

Plus d'infos :

https://www.stat4decision.com/fr/formations/formation-python-pour-la-data-science/

Formation disponible en intra ou en inter-entreprises

PROGRAMME DETAILLE

Jour 1

- Introduction au langage python
  - Pour qui ? Pour quoi faire ? Comment ?
  - Les consoles et le débogage en python
  - Le langage (la syntaxe, les structures (string, listes, dictionnaires…), les opérateurs et les fonctions)
  - Construction d'un outil de calcul de statistiques simples
  - Les classes, les objets et la création de bibliothèques
- Python pour le traitement des données
  - Le processus de développement d'un outil en python pour la data science
  - L'importation des données (texte, csv, Excel, web scrapping., SQL..)
- La préparation et le nettoyage des données (jointure, filtre, transformation, traitement des données manquantes avec pandas, numpy et scipy)

Jour 2 :

- La data visualisation
  - Présentation des principes de la data visualisation
  - Quelques graphiques simples avec matplotlib : scatter plot, box plot, histogrammes…
  - Présentation de visualisation plus avancées
  - Utilisation de Bokeh pour des visualisations interactives
- Introduction au machine learning : présentation des principes de l'apprentissage automatique
- Préparation des données avec la bibliothèque scikit-learn

Jour 3:

- Cas pratiques en machine learning
  - La prévision en marketing
  - L'extraction d'informations pertinentes
  - Les k-means pour la compression d'images
  - Les k-plus proches voisins pour la prévision de scores
  - Le traitement de données textuelles
  - La classification d'image avec le deep learning
- L'automatisation et la mise en production de traitements avec python (pipeline)
- Les bibliothèques pour le big data
  - pyspark
  - Faire des requêtes sur un cluster hadoop
- Discussions et démonstrations de fonctionnalités avancées sur les problématiques des participants