

FR_MUT3476



2 jours en présentiel plus un E-learning préalable sur les bases de Python (1 jour de travail)
Optionnellement et hors formation 1 jour présentiel de travail pratique supplémentaire pourra être organisé directement entre les participants et Safran Analytics.

Blended

DATA BOOTCAMP - Les bases de l'exploration de données v2



OBJECTIFS

Le Data Bootcamp est une formation pratique, où vous apprendrez à explorer en toute autonomie des jeux de données structurés pour en extraire des informations par le biais de la programmation en Python : un outil performant pour l'analyse de données.

En détail, vous serez capable :

- D'approcher méthodologiquement des problèmes via l'analyse de données.
- De nettoyer et transformer un jeu de données dans un format adapté pour le rendre exploitable.
- D'explorer des données diverses à travers la visualisation.
- De juger de la fiabilité des résultats obtenus par le biais de notions statistiques basiques.
- De comprendre les cas d'utilisation du Machine Learning et le Big Data.

...et vous aurez acquis des compétences de base :

- Dans la programmation Python et l'environnement de développement de type Notebook.
- Dans l'utilisation des bibliothèques dédiées à l'analyse de données (Numpy, Pandas, Matplotlib)



PROGRAMME

E-learning

- Introduction à la programmation avec Python
- "Notebooks" pour l'analyse de données
- Introduction à l'exploration de données avec Pandas

Jour 1

Introduction à la data science

- Processus de la data science
- Résolution de problèmes via l'analyse de données
- Applications et cas d'usages

Visualisation de données

- Visualisation Monovariée: Exploration de distributions
- Visualisation Multivariée: Interprétation des relations entre variables
- Analyse des relations entre variables

Manipulation et enrichissement des jeux de données

- Manipulation de données: Sélection et filtrage
- "Feature Engineering": Mieux synthétiser l'information via la création de variables

Jour 2

Transformation et nettoyage de données

- Transformations structurelles de jeux de données
- Agrégation de données pour le calcul de statistiques
- Traitement des problèmes dans la donnée : valeurs manquantes ou aberrantes

Statistiques dans l'analyse de données

- Statistiques et probabilités de base : Variables aléatoires
- P-value et les tests de significativité
- Paradoxes statistiques

Introduction au Big Data et le Machine Learning

- Qu'est-ce que le Big Data? Introduction aux possibilités et les systèmes distribués.
- Le Machine Learning : le principe et les cas d'applications.

Challenge Analyse de données

- Résolution d'une énigme par l'analyse de données.
- Justification des décisions par l'exploration de données

FR_MUT3476



DATA BOOTCAMP - Les bases de l'exploration de données v2 Suite



PUBLIC

Toute personne:

- ayant besoin d'exploiter de la donnée au quotidien de manière fiable et robuste, ou
- avec l'intention d'adresser des problèmes depuis l'angle de la donnée dans un futur proche.



PEDAGOGIE

Formation Action : Alternance d'exercices pratiques de programmation (70%) et apports théoriques (30%)

Partage d'expériences de Data Scientists sur la donnée chez Safran.



SPECIFICITES

- Utilisation fréquente de données pour faire des analyses, indépendamment de l'outil utilisé.
- Prérequis: Avoir des connaissances basiques en Python et la librairie Pandas (E-Learning disponible)