

## MANIPULER, ANALYSER ET VISUALISER DES DONNÉES GRÂCE AUX MODULES PYTHON DE DATA SCIENCE

cette certification vous permet d'attester vos compétences en manipulation, analyse et visualisation des données avec les modules Python de Data Science : Numpy, Pandas, Matplotlib, Seaborn et Plotly.

### Durée

28.00 heures (4.00 jours)

Formation 100% synchrone

### Profils des apprenants

- Professionnels de la donnée (statisticiens, actuaires, data analysts et data scientists) et aux scientifiques voulant attester leurs compétences en manipulation, analyse et visualisation des données avec les modules Python de Data Science

### Accessibilité et délais d'accès

Pour les personnes en situation de handicap, nous étudions les actions que nous pouvons mettre en place pour favoriser leur apprentissage à travers un questionnaire avant formation. Nous nous appuyons également sur un réseau de partenaires locaux.

Pour tout besoin lié à la pédagogie, notre référente est Maud : [maud.hoffmann@axio-formation.com](mailto:maud.hoffmann@axio-formation.com) (également référente handicap)

Pour tout besoin d'ordre administratif, notre référente est Emilie : [emilie.vannieuwenborg@axio-formation.com](mailto:emilie.vannieuwenborg@axio-formation.com)

Pour toute inscription veuillez remplir le formulaire de contact sur notre site. Vous serez recontacté.e. par notre service commercial.

Délai d'accès : 3 semaines

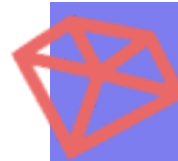
Processus : recueil de besoin, validation prérequis (Entretien et diagnostic initiaux pour adapter le parcours et valider le projet), devis/convention, convocation.

### Qualité et indicateurs de résultats

Taux de présence VS taux d'abandon, taux de satisfaction à chaud et à froid, taux de réussite à l'évaluation finale /taux de certification

### Prérequis

Connaître les fondamentaux de la programmation en langage Python



## OBJECTIFS PÉDAGOGIQUES.

- Comprendre l'écosystème Python de la Data Science en identifiant les principaux environnements, bibliothèques (NumPy, Pandas, Matplotlib, etc.) et leurs usages dans un contexte professionnel.
- Appliquer les techniques de chargement, nettoyage et préparation des données afin de rendre un jeu de données exploitable pour l'analyse (gestion des valeurs manquantes, filtrage, transformation).
- Manipuler des structures de données tabulaires (DataFrame, Series, ndarray) en utilisant les méthodes d'indexation, de tri, de fusion et de transformation pour structurer l'information.
- Analyser des données à l'aide de traitements statistiques et d'opérations vectorisées afin d'extraire des indicateurs pertinents et d'identifier des tendances.
- Évaluer la pertinence des méthodes, outils et traitements utilisés (choix des bibliothèques, types de calculs, structuration des données) au regard d'un besoin métier spécifique.
- Créer des visualisations de données claires, pertinentes et accessibles (graphiques statiques et interactifs) permettant de communiquer efficacement des résultats d'analyse.



### Module 1 : Installer et configurer l'écosystème Python scientifique

- Installation de Python et gestion des versions
- Création et gestion d'un environnement virtuel
- Installation des modules et paquets scientifiques (NumPy, Pandas, Matplotlib, Seaborn, Plotly)
- Prise en main d'un environnement de développement scientifique (Jupyter Notebook, IDE)
- Utilisation d'outils collaboratifs (notebooks partagés, versioning Git)
- Recherche et évaluation de nouvelles librairies selon un besoin métier

### Module 2 : Charger, nettoyer et préparer des données

- Chargement de données (CSV, Excel, JSON...) dans un tableau de données
- Exploration initiale des données (types, valeurs manquantes, structure)
- Nettoyage des données : gestion des valeurs manquantes (remplacement, suppression) et transformation des types de données
- Création et utilisation de masques pour extraire des données
- Filtrage des données : par conditions logiques, par plages temporelles, via expressions rationnelles (regex)
- Rééchantillonnage et changement de fréquence des données

### Module 3 : Manipuler des données tabulaires avec NumPy et Pandas

- Maîtrise des tableaux à n-dimensions (ndarray) et des types de données
- Manipulation des Series et DataFrame
- Indexation des données : indexation implicite (position) et indexation explicite (clé alphanumérique)
- Tri des données selon indices et valeurs
- Transformation des données : transposition, pivot et tableaux croisés dynamiques
- Gestion des index multiples (multi-index)
- Fusion et empilement de données (concaténation, merge, join)
- Structuration des données pour l'analyse métier





### Module 4 : Calculer, analyser et interpréter des données

- Opérations vectorisées pour optimiser les performances
- Application de fonctions sur les données (colonnes/lignes)
- Utilisation des fonctions universelles
- Calculs statistiques descriptifs : moyenne, médiane, min, max et écart-type, centiles
- Création de nouvelles variables (features engineering simple)
- Regroupement de données et agrégation (groupby)
- Calcul de moyennes glissantes
- Utilisation du broadcasting pour les calculs
- Application de fonctions personnalisées et anonymes
- Interprétation des résultats dans un contexte métier

### Module 5 : Visualiser et communiquer des données

- Choix des bibliothèques de visualisation (Matplotlib, Seaborn, Plotly)
- Création de graphiques 2D ( $y = f(x)$ )
- Sélection du type de graphique adapté :
  - histogramme,
  - diagramme en colonnes
  - diagramme circulaire
  - boîte à moustaches
  - carte thermique
- Génération de graphiques à partir de DataFrame
- Création de graphiques interactifs (Plotly)
- Visualisations multi-tracés pour comparaison et corrélation
- Accessibilité des graphiques :
  - titres, labels, légendes
  - lisibilité et contraste
  - adaptation aux personnes en situation de handicap

### Module 6 : Mise en pratique globale et préparation à la certification

- Étude de cas complète : chargement des données, nettoyage et préparation, manipulation et transformation, analyse statistique, visualisation des résultats
- Production d'un code structuré, lisible et optimisé
- Entraînement aux attendus de l'épreuve pratique :
  - utilisation de Pandas / NumPy
  - création de colonnes et traitements
  - filtrage et manipulation
- Révision des notions clés pour l'épreuve théorique (QCM/QCU)



# ORGANISATION DE LA FORMATION.

## Équipe pédagogique

Notre équipe pédagogique maîtrise l'ensemble des sujets proposés à la formation. Nous construisons nos programmes en identifiant les besoins en compétences des futurs apprenants et en collaboration avec nos experts métiers. Axio Formation repose sur une approche personnalisée pour chaque parcours professionnel.

Nous concevons des formations qualifiantes qui non seulement répondent à vos besoins spécifiques, mais vous préparent aussi à exceller dans votre domaine.

Pour tout besoin lié à la pédagogie, notre référente est Maud: [maud.hoffmann@axio-formation.com](mailto:maud.hoffmann@axio-formation.com)

Pour tout besoin d'ordre administratif, notre référente est Emilie: [emilie.vannieuwenborg@axio-formation.com](mailto:emilie.vannieuwenborg@axio-formation.com)

Pour toute inscription veuillez remplir le formulaire de contact sur notre site. Vous serez recontacté.e. par notre service commercial. Délai d'accès : 3 semaines

## Moyens pédagogiques et techniques

- **En présentiel** : Accueil des participants dans une salle dédiée à la formation. Documents supports de formation projetés. Etudes de cas concrets. Quizz et activités collectives en salle; Mise à disposition en ligne de documents supports à la suite de la formation
- **En distanciel** : Classes virtuelles via l'interface Digiforma. Support de formation partagé. Activités d'entraînement en synchrone

## Dispositif de suivi de l'exécution de l'évaluation des résultats de la formation.

- Feuilles d'émergence
- Autoévaluation de niveau en début de formation et fin de formation
- Evaluations d'entraînement tout au long de la formation
- Questionnaire de satisfaction à chaud et à froid
- Réussite à l'évaluation finale (test en ligne du certificateur) ou soutenance devant jury d'évaluation selon certificateur



## POUR LES FORMATIONS ÉLIGIBLES AU CPF.

### Détails sur la certification :

**Titre de la certification** : Manipuler, analyser et visualiser des données grâce aux modules Python de Data Science

**Nom certificateur** : API SOCIETY

**Code** : RS 6763

**Date d'enregistrement** : 01-10-2024

**Détails sur la certification**: Voir fiche EDOF Mon Compte Formation

**Frais d'examen inclus dans le prix des prestations** : 100€/HT par apprenant

En fin de parcours, les apprenants s'engagent à passer l'examen en vue de l'obtention de la certification

Le passage de l'évaluation est obligatoire pour tout candidat utilisant ses droits CPF pour financer la formation. En cas de refus de passer l'examen, un remboursement de la formation pourra lui être demandé.



La certification est composée de deux épreuves :

- une épreuve théorique durant laquelle le candidat doit répondre à un questionnaire de type QCM portant sur les modules Python de Data Science
- une épreuve pratique durant laquelle le candidat doit utiliser les modules Python de Data Science au travers d'exercices de programmation Thèmes abordés : écosystème Python scientifique, Numpy, Pandas, Matplotlib, Seaborn et Plotly.