

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.

Points forts de la formation: Cours individuels ou en mini groupe limité à 10 personnes pour plus d'interactivité, Préparation d'une certification officielle

Durée

28.00 heures (4.00 jours)

Formation 100% synchrone en présentiel ou en classe virtuelle selon la session sélectionnée sur Mon Compte Formation.

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 référent est Patrick : gestion@axio-academy.com (également référent handicap)

Pour tout besoin d'ordre administratif, notre référente est Emilie : gestion@axio-academy.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, 11 jours ouvrés en financement CPF

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

Pour les inscriptions réalisées via Mon Compte Formation, les informations précontractuelles complètes relatives à la session, au prix, aux dates, aux horaires, aux modalités d'inscription, aux conditions de financement, aux modalités de certification et aux éventuels frais annexes sont accessibles sur la fiche EDOF correspondante avant validation de la commande.

L'inscription définitive intervient exclusivement après validation de la commande par le titulaire sur la plateforme Mon Compte Formation.

Prérequis

Connaître les fondamentaux de la programmation en langage Python

En cas de séance en distanciel, il faut également savoir utiliser une plateforme collaborative en ligne (visio-conférence)



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

OBJECTIFS PÉDAGOGIQUES.

- Installer et configurer un environnement Python dédié à la Data Science, en utilisant notamment Jupyter Notebook, un IDE adapté et Git pour le versioning et le travail collaboratif.
- Importer, nettoyer et préparer des jeux de données avec Pandas et NumPy, en traitant les valeurs manquantes, en modifiant les types de données et en appliquant des filtres, masques, expressions régulières et rééchantillonnages.
- Manipuler et structurer des données tabulaires avec NumPy et Pandas, à partir des objets ndarray, Series et DataFrame, en utilisant l'indexation, le tri, la transposition, les pivots, les MultiIndex ainsi que les opérations de fusion et d'empilement.
- Calculer, analyser et interpréter des données avec NumPy et Pandas, en mobilisant les fonctions statistiques, les opérations vectorisées, le broadcasting, les agrégations, les fonctions universelles et les fonctions personnalisées.
- Évaluer la pertinence des bibliothèques et méthodes utilisées — notamment NumPy, Pandas, Matplotlib, Seaborn et Plotly — au regard d'un besoin d'analyse ou de traitement de données spécifique.
- Créer des visualisations statiques, statistiques et interactives avec Matplotlib, Seaborn et Plotly, en choisissant le type de graphique adapté et en appliquant les recommandations d'accessibilité, de lisibilité et de restitution des résultats.

CONTENU DE LA FORMATION.



1 : Mettre en place son environnement Python pour la Data Science

- Installer Python et comprendre la gestion de ses différentes versions.
- Créer et administrer un environnement virtuel dédié à un projet Data Science.
- Installer et mettre à jour les principales bibliothèques scientifiques : NumPy, Pandas, Matplotlib, Seaborn et Plotly.
- Prendre en main un environnement de développement scientifique tel que Jupyter Notebook ou un IDE Python.
- Utiliser des outils collaboratifs pour partager et versionner les travaux, notamment les notebooks et Git.
- Rechercher de nouvelles bibliothèques Python et évaluer leur pertinence au regard d'un besoin professionnel.
- Exercice d'application : créer un environnement Python opérationnel, installer NumPy, Pandas, Matplotlib, Seaborn et Plotly, ouvrir un notebook sous Jupyter Notebook et sélectionner une bibliothèque complémentaire en justifiant son intérêt.

2 : Importer, nettoyer et préparer les données avec Pandas et NumPy

- Importer différents formats de données avec Pandas : CSV, Excel, JSON.
- Explorer la structure d'un jeu de données à l'aide des objets DataFrame et Series.
- Identifier les types de variables et repérer les valeurs manquantes.
- Nettoyer les données avec Pandas : suppression, remplacement ou traitement des valeurs manquantes.
- Modifier les types de données lorsque leur format n'est pas adapté.
- Concevoir des masques avec NumPy et Pandas afin d'extraire certaines observations.
- Filtrer les données à partir de conditions logiques.
- Réaliser des filtrages selon des plages temporelles.
- Utiliser les expressions rationnelles (regex) pour rechercher et extraire des données correspondant à un motif.
- Modifier la fréquence d'échantillonnage des données avec les fonctionnalités de Pandas.
- Exercice d'application : charger dans Pandas un fichier comportant des données incomplètes, identifier et traiter les valeurs manquantes, construire des masques, réaliser plusieurs filtres et préparer un DataFrame exploitable pour l'analyse.

3 : Structurer et transformer des données avec NumPy et Pandas

- Manipuler les tableaux multidimensionnels ndarray avec NumPy.
- Comprendre et gérer les principaux types de données.
- Créer et manipuler des Series et DataFrame avec Pandas.
- Utiliser l'indexation implicite par position.



- Exploiter l'indexation explicite par clé alphanumérique ou temporelle.
- Trier les données selon leurs index ou leurs valeurs.
- Transposer les données avec NumPy et Pandas.
- Construire des tableaux croisés et effectuer des opérations de pivot avec Pandas.
- Créer et exploiter des MultiIndex.
- Regrouper plusieurs sources grâce aux opérations concat, merge et join de Pandas.
- Structurer les données afin de préparer leur analyse dans un contexte professionnel.
- Exercice d'application : à partir de plusieurs DataFrame Pandas, modifier les types de données, créer des index, trier et transposer les tableaux, puis fusionner plusieurs jeux de données à l'aide de concat, merge ou join.



4 : Calculer, analyser et interpréter des données avec NumPy et Pandas

- Utiliser les opérations vectorisées de NumPy et Pandas pour optimiser les traitements.
- Appliquer des fonctions et méthodes sur les lignes et colonnes d'un DataFrame.
- Mobiliser les fonctions universelles NumPy (ufuncs).
- Calculer des statistiques descriptives avec NumPy et Pandas : moyenne, médiane, minimum, maximum, centiles et écart-type.
- Créer de nouvelles variables à partir de colonnes existantes.
- Regrouper et agréger les données avec la fonction groupby de Pandas.
- Calculer des moyennes glissantes avec les fonctions rolling de Pandas.
- Utiliser les règles de broadcasting NumPy pour effectuer des opérations sur des tableaux.
- Appliquer des fonctions personnalisées et des fonctions anonymes lambda aux données.
- Interpréter les résultats obtenus au regard d'un besoin métier.
- Exercice d'application : analyser un jeu de données avec NumPy et Pandas, créer de nouvelles colonnes, réaliser des agrégations avec groupby, calculer des statistiques et des moyennes glissantes, puis mobiliser le broadcasting pour optimiser certains calculs.

5 : Visualiser et communiquer les données avec Matplotlib, Seaborn et Plotly

- Identifier les spécificités des bibliothèques Matplotlib, Seaborn et Plotly.
- Choisir la bibliothèque de visualisation la plus adaptée au besoin.
- Créer des graphiques 2D avec Matplotlib.
- Générer différents types de représentations : histogrammes, diagrammes en colonnes, diagrammes circulaires, boîtes à moustaches et cartes thermiques.
- Générer directement des graphiques à partir de DataFrame Pandas.
- Utiliser Seaborn pour produire des visualisations statistiques.
- Créer des visualisations interactives avec Plotly.
- Concevoir des graphiques comportant plusieurs tracés afin de comparer plusieurs séries ou identifier des corrélations.
- Personnaliser titres, axes, labels et légendes.
- Améliorer l'accessibilité des graphiques : contraste, taille et lisibilité des textes, épaisseur des lignes et bordures, légendes explicites.
- Structurer une restitution visuelle facilitant la compréhension des résultats par différents utilisateurs.
- Exercice d'application : à partir d'un même jeu de données, créer plusieurs représentations avec Matplotlib, Seaborn et Plotly, comparer les résultats, retenir la visualisation la plus pertinente puis l'adapter afin de respecter les recommandations d'accessibilité.



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 Academy Impulseo repose sur une approche personnalisée pour chaque parcours professionnel. Axio Academy Impulseo accompagne le candidat dans la compréhension des attendus de certification et dans l'organisation du passage de l'évaluation, selon les modalités prévues par le certificateur et précisées sur la fiche EDOF.

La formation est animée par un formateur expert en programmation Python, algorithmique, automatisation, traitement de données et développement d'applications professionnelles. Nous concevons des formations certifiantes qui non seulement répondent à vos besoins spécifiques, mais vous préparent aussi à exceller dans votre domaine.

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
- Un support technique est également mis en place pour les séances en distanciel.

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

- Feuilles d'émargement
- Autoévaluation de niveau en début de formation et fin de formation
- Exercices d'entraînement tout au long de la formation
- Questionnaire de satisfaction à chaud et à froid
- Réussite à l'évaluation finale prévue par le certificateur
- Contrôle permanent des acquis par le formateur

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

Financements possibles : CPF, OPCO, Fonds propres, France Travail et Autres financements possibles (nous contacter).

Accompagnement Axio : information, préparation, inscription, convocation, calendrier, relance, suivi du passage.

Le passage de l'évaluation certificative est **obligatoire pour tout candidat mobilisant ses droits CPF**. En cas d'absence ou de refus de présentation à l'évaluation, cette situation est tracée dans le dossier du bénéficiaire et peut avoir une incidence sur la validation de la certification, selon les règles applicables de Mon Compte Formation et du certificateur.

Les modalités complètes de certification, les conditions de passage, les éventuelles dates prévisionnelles, le coût de certification et les informations de session sont précisés sur la fiche EDOF accessible avant validation de la commande.



Organisées par l'organisme certificateur, les épreuves de la certification sont composées d'un questionnaire théorique évaluant les connaissances sur les modules Python de Data Science, et d'exercices pratiques évaluant les compétences de programmation du candidat.

A l'aide du logiciel API Cert permettant des captations audio vidéo durant les épreuves d'évaluation, cette certification est passable à distance et en asynchrone.

Le jury est composé d'un surveillant qui garantit l'identité du candidat et l'absence de fraudes grâce aux captations, d'un professionnel de la Data Science avec Python certifié qui corrige les exercices programmés par le candidat sans accéder aux captations afin de préserver l'anonymat des candidats et l'équité de la correction, et d'un président qui attribue ou non la certification selon les comptes rendus du surveillant et du correcteur.