

Intitulé de la certification

Manipuler, analyser et visualiser des données grâce aux modules Python de Data Science

Description des métiers

Cette certification s'adresse aux 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

Référentiel d'activités, de compétences et d'évaluation

Référentiel d'activités décrit les situations de travail et les activités exercées, les métiers ou emplois visés	Référentiel de compétences identifie les connaissances et les compétences, y compris transversales, qui découlent du référentiel d'activités	Référentiel d'évaluation définit les critères et les modalités d'évaluation des acquis	
		Modalités d'évaluation	Critères d'évaluation
<p>A1 - Installer l'écosystème Python scientifique</p> <p>(emploi / métier : exemple de situation de travail)</p> <p><u>Data Scientist, Scientifiques tous secteurs</u></p>	<p>C1 - Installer dans un environnement virtuel des modules et paquets scientifiques pour manipuler des données</p> <p>C2 - Utiliser un environnement de développement scientifique afin de bénéficier d'une solution intégrée et ainsi être plus rapide et performant</p> <p>C3 - Utiliser des outils collaboratifs afin de travailler à plusieurs data scientists sur les mêmes données</p> <p>C4 - Trouver de nouvelles librairies et juger de leur pertinence en vue de répondre à un besoin particulier</p>	<p>Épreuve théorique : le candidat doit répondre à un questionnaire (QCU, QCM et informations à saisir) de 40 questions tirées aléatoirement parmi une banque de plus de 200 questions</p> <p>Durée totale de l'épreuve théorique : 20 minutes</p>	<p>Épreuve théorique : le candidat a répondu 60% de bonnes réponses ou plus au questionnaire, soit plus de 24 réponses justes sur 40 questions posées</p>
<p>A2 - Charger, nettoyer et préparer des données</p> <p>(emploi / métier : exemple de situation de travail)</p> <p><u>Data Scientist</u> : charger un fichier CSV, indexer une colonne contenant une date, remplacer les valeurs nulles par 0</p>	<p>C5 - Charger différents formats de données dans un tableau de données indexées ou non en vue de leur appliquer des traitements</p> <p>C6 - Appliquer des fonctions et méthodes sur un tableau de données afin de le nettoyer et de le préparer à un futur traitement</p> <p>C7 - Concevoir des masques pour extraire des données</p> <p>C8 - Filtrer des données selon des plages temporelles afin d'extraire les données correspondantes à un moment précis</p> <p>C9 - Filtrer des données à l'aide d'expressions rationnelles (regex) afin d'extraire les données correspondantes à un motif précis</p> <p>C10 - Changer la fréquence d'échantillonnage de données afin de les ramener sur une échelle adaptée au besoin</p>	<p>Épreuve théorique : le candidat doit répondre à un questionnaire (QCU, QCM et informations à saisir) de 40 questions tirées aléatoirement parmi une banque de plus de 200 questions</p> <p>Durée totale de l'épreuve théorique : 20 minutes</p> <p>Épreuve pratique : en utilisant la bibliothèque NumPy ou Pandas, le candidat doit savoir remplacer les données manquantes d'un tableau, filtrer des données avec des masques, et leur appliquer un traitement simple</p> <p>Durée totale de l'épreuve pratique : 2 heures</p> <p>Temps alloué à cet exercice : 15 minutes</p>	<p>Épreuve théorique : le candidat a répondu 60% de bonnes réponses ou plus au questionnaire, soit plus de 24 réponses justes sur 40 questions posées</p> <p>Épreuve pratique : le travail rendu par le candidat est corrigé par un membre du jury professionnel Data Scientist Python en activité selon les critères suivants :</p> <p>CR1 - le candidat sait identifier les données manquantes</p> <p>CR2 - le candidat sait traiter les données manquantes</p> <p>CR3 - le candidat est capable d'utiliser les masques pour filtrer des données</p>

			<p>CR4 - le candidat sait appliquer un traitement simple aux données</p> <p>CR5 - le code écrit par le candidat est simple et élégant</p>
<p>A3 - Travailler des données en mode tableau</p> <p>(emploi / métier : exemple de situation de travail)</p> <p><u>Océanographe</u> : empiler et fusionner des données pour rassembler les données de plusieurs capteurs dans un seul tableau</p>	<p>C7 - Concevoir des masques pour extraire des données</p> <p>C8 - Filtrer des données selon des plages temporelles afin d'extraire les données correspondantes à un moment précis</p> <p>C9 - Filtrer des données à l'aide d'expressions rationnelles (regex) afin d'extraire les données correspondantes à un motif précis</p> <p>C10 - Changer la fréquence d'échantillonnage de données afin de les ramener sur une échelle adaptée au besoin</p> <p>C11 - Maîtriser les tableaux à n-dimensions (ndarray) et les types de données (entier, réel, chaîne de caractères, date et heure) afin de minimiser l'utilisation de la mémoire</p> <p>C12 - Manipuler les tableaux à données indexées à une dimension (Serie) et à deux dimensions (DataFrame) afin d'accéder aux fonctionnalités offertes par le module scientifique</p> <p>C13 - Maîtriser l'indexation implicite (indilage par position numérique) afin d'extraire des données et/ou de leur appliquer un traitement</p> <p>C14 - Maîtriser l'indexation explicite (indilage par clé alphanumérique) afin de manipuler plus facilement des données avec des identifiants temporels ou alphanumériques</p> <p>C15 - Faire pivoter des données pour obtenir un tableau croisé dynamique</p> <p>C16 - Maîtriser les index multiples (multi-index) afin d'accéder plus facilement aux données et/ou de simuler des dimensions supplémentaires</p> <p>C17 - Empiler et fusionner des tableaux afin de regrouper des données</p> <p>C18 - Transposer des données (inversion ligne / colonne) pour en faciliter l'analyse, ou les préparer en vue de les empiler ou les fusionner</p>	<p>Épreuve théorique : le candidat doit répondre à un questionnaire (QCU, QCM et informations à saisir) de 40 questions tirées aléatoirement parmi une banque de plus de 200 questions</p> <p>Durée totale de l'épreuve théorique : 20 minutes</p> <p>Épreuve pratique : en utilisant la bibliothèque Pandas ou Numpy, le candidat doit savoir changer le type de données d'une colonne d'un tableau, indexer une colonne, transposer et trier des données</p> <p>Durée totale de l'épreuve pratique : 2 heures</p> <p>Temps alloué à cet exercice : 30 minutes</p>	<p>Épreuve théorique : le candidat a répondu 60% de bonnes réponses ou plus au questionnaire, soit plus de 24 réponses justes sur 40 questions posées</p> <p>Épreuve pratique : le travail rendu par le candidat est corrigé par un membre du jury professionnel Data Scientist Python en activité selon les critères suivants :</p> <p>CR6 - le candidat est capable de changer le type de données de la colonne des données d'un tableau</p> <p>CR7 - le candidat sait indexer les données d'un tableau</p> <p>CR8 - le candidat est capable de transposer un tableau</p> <p>CR9 - le candidat sait trier des données d'un tableau</p> <p>CR10 - le code écrit par le candidat est simple et élégant</p>

	<p>C19 - Trier des données en fonction des indices ou des valeurs du tableau en vue de les exploiter par ordre alphabétique, par nombre croissant ou décroissant, par date</p>		
<p>A4 - Calculer, analyser et interpréter des données statistiques</p> <p>(emploi / métier : exemple de situation de travail)</p> <p>Trader : appliquer des fonctions statistiques à des données afin d'évaluer le risque sur une position</p>	<p>C7 - Concevoir des masques pour extraire des données</p> <p>C8 - Filtrer des données selon des plages temporelles afin d'extraire les données correspondantes à un moment précis</p> <p>C9 - Filtrer des données à l'aide d'expressions rationnelles (regex) afin d'extraire les données correspondantes à un motif précis</p> <p>C10 - Changer la fréquence d'échantillonnage de données afin de les ramener sur une échelle adaptée au besoin</p> <p>C11 - Maîtriser les tableaux à n-dimensions (ndarray) et les types de données (entier, réel, chaîne de caractères, date et heure) afin de minimiser l'utilisation de la mémoire</p> <p>C12 - Manipuler les tableaux à données indexées à une dimension (Serie) et à deux dimensions (DataFrame) afin d'accéder aux fonctionnalités offertes par le module scientifique</p> <p>C13 - Maîtriser l'indexation implicite (indilage par position numérique) afin d'extraire des données et/ou de leur appliquer un traitement</p> <p>C14 - Maîtriser l'indexation explicite (indilage par clé alphanumérique) afin de manipuler plus facilement des données avec des identifiants temporels ou alphanumériques</p> <p>C19 - Trier des données en fonction des indices ou des valeurs du tableau en vue de les exploiter par ordre alphabétique, par nombre croissant ou décroissant, par date</p> <p>C20 - Connaître et utiliser les opérations vectorisées pour gagner en performance</p> <p>C21 - Appliquer des fonctions et méthodes sur une ou plusieurs colonnes ou lignes d'un tableau de données afin d'appliquer des traitements mathématiques (nombre, minimum, maximum, moyenne, médiane, centile, écart-type)</p> <p>C22 - Utiliser les fonctions universelles selon une dimension</p>	<p>Épreuve théorique : le candidat doit répondre à un questionnaire (QCU, QCM et informations à saisir) de 40 questions tirées aléatoirement parmi une banque de plus de 200 questions</p> <p>Durée totale de l'épreuve théorique : 20 minutes</p> <p>Épreuve pratique : en utilisant la bibliothèque Pandas ou Numpy, le candidat doit savoir créer une nouvelle colonne en appliquant une opération sur une colonne existante, et appliquer un traitement mathématique simple sur une colonne</p> <p>Durée totale de l'épreuve pratique : 2 heures</p> <p>Temps alloué à cet exercice : 30 minutes</p>	<p>Épreuve théorique : le candidat a répondu 60% de bonnes réponses ou plus au questionnaire, soit plus de 24 réponses justes sur 40 questions posées</p> <p>Épreuve pratique : le travail rendu par le candidat est corrigé par un membre du jury professionnel Data Scientist Python en activité selon les critères suivants :</p> <p>CR11 - le candidat sait produire une nouvelle colonne résultant d'une opération sur une autre colonne ou d'un calcul simple entre deux colonnes</p> <p>CR12 - le candidat sait utiliser le broadcasting pour réaliser des opérations sur le tableau de données</p> <p>CR13 - le candidat sait faire appel à des fonctions universelles pour réaliser des opérations sur une partie ou l'ensemble des données</p> <p>CR14 - le candidat sait appliquer des opérations vectorisées afin de réaliser des opérations performantes en temps de calcul</p> <p>CR15 - le candidat est capable d'appliquer un traitement mathématique simple sur une colonne</p> <p>CR16 - le code écrit par le candidat est simple et élégant</p>

	<p>C23 - Regrouper des données en fonction de la valeur de certaines colonnes pour appliquer un calcul</p> <p>C24 - Calculer des moyennes glissantes (lissage des données) en vue de les visualiser</p> <p>C25 - Utiliser les règles de broadcasting pour effectuer des calculs</p> <p>C26 - Appliquer une fonction ou une fonction anonyme aux données pour réaliser des analyses spécifiques</p>		
<p>A5 - Représenter graphiquement des données</p> <p>(emploi / métier : exemple de situation de travail)</p> <p><u>Prévisionniste Météo</u> : représenter visuellement des données météorologiques pour alerter sur le dérèglement climatique</p>	<p>C27 - Choisir le bon module ou paquet de visualisation des données afin de répondre au besoin exprimé</p> <p>C28 - Afficher un graphique en 2 dimensions de type $y = f(x)$ à partir d'un jeu de données simple</p> <p>C29 - Identifier le bon type de graphique (histogramme, graphique en colonne, diagramme circulaire, boîte à moustaches, carte thermique ...) en fonction des données à visualiser afin de générer un graphique pertinent</p> <p>C30 - Connaître et utiliser les fonctions et méthodes des tableaux de données indexées afin de générer très facilement des graphiques</p> <p>C31 - Générer des graphiques interactifs afin de permettre à l'utilisateur d'interagir avec le graphique (isolation de données, agrandissement d'une zone, variation interactive des données)</p> <p>C32 - Concevoir des graphiques contenant plusieurs tracés afin de regrouper des données différentes, de montrer des corrélations ou des différences, ou de synthétiser une représentation des données</p> <p>C33 - Intégrer les recommandations d'amélioration de l'accessibilité (via l'ajout de titres, étiquettes, labels et légendes, l'amélioration de la lisibilité des textes, l'épaississement des lignes et des bordures, le choix de couleurs contrastées) afin de générer des graphiques accessibles à une personne en situation de handicap</p>	<p>Épreuve théorique : le candidat doit répondre à un questionnaire (QCU, QCM et informations à saisir) de 40 questions tirées aléatoirement parmi une banque de plus de 200 questions</p> <p>Durée totale de l'épreuve théorique : 20 minutes</p> <p>Épreuve pratique : le candidat doit adapter des exemples de graphique créés avec les bibliothèques Matplotlib, Plotly et Seaborn (courbe, diagramme en bâton et histogramme) aux données d'un tableau fourni par l'énoncé en respectant les recommandations d'amélioration de l'accessibilité</p> <p>Durée totale de l'épreuve pratique : 2 heures</p> <p>Temps alloué à cet exercice : 45 minutes</p>	<p>Épreuve théorique : le candidat a répondu 60% de bonnes réponses ou plus au questionnaire, soit plus de 24 réponses justes sur 40 questions posées</p> <p>Épreuve pratique : le travail rendu par le candidat est corrigé par un membre du jury professionnel Data Scientist Python en activité selon les critères suivants :</p> <p>CR17 - le candidat sait adapter un exemple de la bibliothèque Matplotlib</p> <p>CR18 - le candidat est capable d'adapter un exemple de la bibliothèque Plotly</p> <p>CR19 - le candidat sait modifier un exemple de la bibliothèque Seaborn</p> <p>CR20 - le candidat sait créer un graphique adapté à une personne en situation de handicap</p> <p>CR21 - le code écrit par le candidat est simple et élégant</p>