

Référentiel d'activités, de compétences et d'évaluation

REFERENTIEL D'ACTIVITES <i>décrit les situations de travail et les activités exercées, les métiers ou emplois visés</i>	REFERENTIEL DE COMPETENCES <i>identifie les compétences et les connaissances, y compris transversales, qui découlent du référentiel d'activités</i>	REFERENTIEL D'ÉVALUATION <i>définit les critères et les modalités d'évaluation des acquis</i>	
		MODALITÉS D'ÉVALUATION	CRITÈRES D'ÉVALUATION
Bloc n°1 Collecte des données : exploration et requête des différents types de bases de données, et récupération des données			
<p>A1.1 Identification des données disponibles</p> <p>A1.2 Recherche de données à partir de besoins définis</p> <p>A1.3 Gestion d'opérations de collecte et de qualification de données</p>	<p>C1.1 Explorer le contenu de bases de données afin d'identifier les possibilités d'utilisation et d'interprétation des données</p> <p>C1.2 Modéliser des bases de données relationnelles (notamment SQL) afin de répondre aux besoins des utilisateurs</p> <p>C1.3 Réaliser des requêtes sur des bases de données relationnelles pour extraire des données</p> <p>C1.4 Réaliser des agrégations, des jointures et des sous-requêtes sur des bases de données (notamment SQL) afin d'exploiter les données collectées de manière avancée</p> <p>C1.5 Automatiser des collectes de données afin d'exploiter les contenus et les informations récoltés sur des pages web ("web scraping")</p>	<p><u>Mise en situation professionnelle :</u> Mise en place d'une collecte de données</p> <p><u>Cadre de mise en œuvre et de déroulement de l'évaluation</u> Les candidats sont évalués lors de la réalisation d'une mise en situation au cours de laquelle ils doivent effectuer des recherches à partir de questions proposées, et automatiser la récupération des données depuis une base de données.</p>	<ul style="list-style-type: none"> - Une exploration de la base est effectuée pour analyser la qualité de données - La modélisation de la base de données est correctement interprétée par le candidat - Des requêtes sont effectuées sur la base de données relationnelle - Des agrégations, des jointures et des sous-requêtes sur la base de données sont réalisées

	<p>C1.6 Automatiser des requêtes à une "API REST" afin de partager des données entre différentes applications</p> <p>C1.7 Utiliser des outils adaptés afin de traduire des adresses postales en coordonnées géographiques (latitude/longitude) et inversement</p> <p>C1.8 Contrôler des modalités de collecte et d'utilisation de données et mesurer les enjeux du RGPD afin d'estimer les risques et les responsabilités de chacun</p>	<p>Evaluation individuelle des candidats par un examinateur</p>	<ul style="list-style-type: none"> - Des données sont collectées automatiquement depuis des pages web grâce au webscraping - Des requêtes API REST sont utilisées et automatisées pour compléter les données de la base de données - Des adresses postales sont traduites en coordonnées latitude/longitude et inversement - Les enjeux et les responsabilités RGPD sont précisés
--	---	---	---

Bloc n°2 Automatisation du traitement des données : nettoyage, complétion, correction, uniformisation

<p>A2.1 Analyse des besoins de traitement de données</p> <p>A2.2 Structuration d'outils et d'algorithmes de traitements de données</p>	<p>C2.1 Utiliser un logiciel adapté (par exemple Google colab, Spyder ou Jupyterlab) afin de rédiger et exécuter un script pour traiter des données à partir d'objectifs définis</p> <p>C2.2 Utiliser des algorithmes en langage Python (variables, boucles, itérateur, conditions, fonctions) afin de traduire des</p>	<p><u>Mise en situation professionnelle</u></p> <p>Structuration d'outils /programmes de traitement de données selon des objectifs</p>	<ul style="list-style-type: none"> - Un logiciel de rédaction et d'exécution de code (IDE) est utilisé correctement - L'algorithmie utilisée en langage Python est
--	---	---	--

<p>A2.3 Fiabilisation d'outils de traitements de bases de données et organisation de codes</p>	<p>besoins de traitements de données en fonction des objectifs d'exploitation C2.3 Manipuler des structures de données (chaînes de caractères, listes, dictionnaires et tuples) afin de permettre leur exploitation à travers des applications C2.4 Utiliser la programmation orientée objet en Python (création de classes et d'héritage) afin de faciliter la maintenance d'une application C2.5 Effectuer des calculs matriciels (notamment avec NumPy) afin de transformer des données dans le but de les rendre exploitables C2.6 Utiliser les DataFrames (notamment avec Pandas) afin de faciliter l'import, la manipulation et la fusion de données C2.7 Retraiter les valeurs aberrantes (outliers) et les valeurs manquantes afin d'éviter les impacts sur l'exploitation et le traitement des bases de données C2.8 Utiliser les expressions régulières (RegEx) pour traiter les valeurs textuelles C2.9 Utiliser les principes du "clean code" afin d'organiser un code Python</p>	<p><u>Cadre de mise en œuvre et de déroulement de l'évaluation</u> Les candidats doivent, à partir d'objectifs définis de traitement de données, structurer des outils et utiliser des algorithmes afin de manipuler des tableaux de données, et automatiser le nettoyage notamment des valeurs manquantes et aberrantes.</p> <p>Evaluation individuelle des candidats par un examinateur</p>	<p>adaptée aux bonnes pratiques</p> <ul style="list-style-type: none"> - Les structures de données utilisées (chaînes de caractères, listes, dictionnaires et tuples) sont pertinentes - Le programme instancie des objets qui héritent des propriétés (méthodes et attributs) de leur classe grâce à la programmation orientée objet en Python - Les retraitements de données par le calcul matriciel sont conformes aux attentes du client/utilisateur - Des DataFrames Pandas sont utilisés pour la manipulation des données - Le nettoyage des valeurs manquantes et aberrantes est automatisé - Les expressions régulières (RegEx) sont utilisées correctement sur du texte
--	--	---	--

			- Les principes du "clean code" sont appliqués, notamment l'indentation, le nommage des variables et les commentaires
--	--	--	---

Bloc n°3 Modélisation des données structurées : identification des corrélations existantes et utilisation du Machine Learning pour établir des prévisions			
<p>A3.1 Sélection d'informations utiles</p> <p>A3.2 Elaboration de structures de traitement d'informations</p> <p>A3.3 Modélisation de données</p>	<p>C3.1 Maîtriser le process de Machine Learning (syntaxe scikit-learn, train-test-split, fit, predict, score) afin de permettre à des algorithmes d'apprendre automatiquement à partir de données</p> <p>C3.2 Modéliser des régressions supervisées et interpréter les métriques associées afin de définir des modèles de prévisions</p> <p>C3.3 Modéliser des classifications supervisées et interpréter les métriques associées afin de définir des modèles prédictifs</p> <p>C3.4 Utiliser le clustering non-supervisé et ses métriques afin de révéler des similarités entre groupes homogènes de données</p> <p>C3.5 Utiliser des outils et des méthodes statistiques afin de réaliser des réductions de dimension non-supervisée en Machine Learning (ACP, scikit-learn)</p> <p>C3.6 Traiter automatiquement le langage naturel (NLP) à partir de texte brut afin d'en tirer de la valeur en fonction de classification (Sentiment analysis)</p>	<p><u>Mise en situation professionnelle</u> Modélisation de données structurées et détermination de prévisions</p> <p><u>Descriptif de l'évaluation</u> Les candidats doivent présenter une proposition de modélisation de données structurées permettant de décrire les données de manière simple, d'en tirer des tendances et de prévoir des valeurs futures.</p> <p>Evaluation individuelle des candidats par un examinateur</p>	<ul style="list-style-type: none"> - Le process de Machine Learning est utilisé et intégré à la modélisation - Les régressions supervisées sont modélisées et les métriques associées sont interprétées correctement - Les classifications supervisées sont modélisées et les métriques associées sont interprétées correctement - Le clustering non-supervisé et ses métriques sont interprétés correctement et permettent de révéler des similarités - La réduction de dimension non-supervisée en Machine Learning (ACP,

	<p>C3.7 Utiliser les statistiques descriptives (variance, quantiles, coefficients de corrélation) afin de modéliser les données et en faire émerger des informations pertinentes</p> <p>C3.8 Utiliser le théorème central limite pour standardiser des données</p>		<p>scikit-learn) est utilisée pour réduire la taille du modèle</p> <ul style="list-style-type: none"> - Le langage naturel (NLP) est traité automatiquement pour en tirer de la valeur (scikit-learn, NLTK) - Les statistiques descriptives (variance, quantiles, coefficients de corrélation) sont utilisées pour expliquer les données - Le théorème central limite est utilisé pour la standardisation des données
--	--	--	--

Bloc n°4 Visualisation des données : valorisation et interprétation des données pertinentes, et mise en forme dans un tableau de bord			
<p>A4.1 Visualisation de données</p> <p>A4.2 Présentation et partage de données et d'informations</p>	<p>C4.1 Prioriser et identifier des données et des informations à rendre accessibles et à présenter visuellement afin de structurer des représentations graphiques de tableaux bord</p> <p>C4.2 Utiliser les visualisations descriptives, notamment les nuages de points, les boîtes</p>	<p><u>Mise en situation professionnelle</u></p> <p>Présentation d'un Tableau de bord réalisé à partir de besoins d'un client</p> <p><u>Cadre de mise en œuvre et de déroulement de l'évaluation</u></p>	<ul style="list-style-type: none"> - Les besoins et attentes concernant le tableau de bord sont respectés - Les visualisations descriptives sont correctement utilisées pour décrire les données

<p>à moustache et les histogrammes afin de représenter graphiquement des données statistiques et des informations modélisées à destination d'analyste de données</p> <p>C4.3 Manipuler la Dataviz interactive et dynamique (par exemple avec Plotly ou Bokeh) afin de réaliser différents types de représentations graphiques et visuelles de données à destination d'utilisateurs opérationnels</p> <p>C4.4 Réaliser de la cartographie (notamment avec Folium) afin de représenter des informations géographiques</p> <p>C4.5 Utiliser un Tableur, et notamment les tableaux croisés dynamiques afin de proposer des croisements de variables pour obtenir des informations recherchées</p> <p>C4.6 Réaliser des tableaux de bord avec des outils de Business Intelligence (par exemple PowerBI ou Tableau) afin d'intégrer et de croiser des informations utiles à des approches stratégiques de problématiques</p>	<p>Les candidats doivent présenter un tableau de bord réalisé à partir d'un cahier des charges détaillant, à partir de bases de données existantes, des objectifs et des attentes en termes de présentations graphiques et visuelles d'informations et de données croisées.</p> <p>Évaluation individuelle des candidats par un examinateur.</p>	<ul style="list-style-type: none"> - Les visualisations interactives et dynamiques (Plotly ou Bokeh) sont utilisées correctement - Une cartographie est présente - Les Tableurs et tableaux croisés dynamiques sont utilisés pour croiser des variables - Un outil de Business Intelligence (PowerBI ou Tableau) permet une utilisation stratégique du tableau de bord
--	--	--